

Db	4731	CCGCTGTTTGGAGAACCATGTCGGCTGTGACAAAGAAATGGGGAGTGGACACAGACAGACC	4730
Qy	1197	caaccagggccgclgtlaactgcttgcgcgaagtagacacttggaga1tggaaagtctgcgct	1256
Db	4791	CAACCAAGCGCTGCTTACAGTGTGGCACATACACTGAGATGGAAAGTCTGGACACT	4850
Qy	1257	latcaatgctctgcttaacgaacaatgtgcgtctgcagtlccatttgcctcttgaactaac	1316
Db	4851	CATCAATGCTGCTTAACTAAATAATGGCGGCTGTAGTAAATTTGCCATCTCCACACAC	4910
Qy	1317	tgaaagaat1caaaagatagtatctccgaaagccagaactacaacgggttggtaatcgtctg	1376
Db	4911	TGGGCAAGTAAAGAAAGCACTTGTACTTGCAACGCCAAACTACATTTGAGAGTGGCATTTACTG	4970
Qy	1377	ccggggcagcatctcaacgggaggtctcccaagaaccttcgaactcccaagtaacttccca	1436
Db	4971	CCGGCGCAGCATTTATCAGAGCGTTCGCCAAGAAACCGAATACTCCAGTATTTCCTTCCA	5030
Qy	1437	gttgcagagagcatgtcgtlcccgaaagcttgc1tggaaecttgcccttcaccgtctgcgc	1496
Db	5031	GTTTCAGAGACCATTTGCTGTGAAGATCGTGGCGCCAGGCCCTTCACGTGTTTGTGACC	5090
Qy	1497	tttctcagctctcttaatcatbaagccccggat1taaagaattggatctgaagggccatcat	1556
Db	5091	TTTTATCTCGACGCTTTGATGAGGAGAGCTGGGGTTAAAGACTGGGCAAAATACGGGTTTAA	5150
Qy	1557	gtcccaagctctctcgtcatcaacgttgg1tggccgtccgaacagctgcgcgtt1tggaaacctaa	1616
Db	5151	GGCCGACAGTCTTTCGGTACCATGTGGTGGCTGGCCACACAGCTGCTTCTGGAAACCTGAA	5210
Qy	1617	agtgaaccaagt1tgcacgaaccttccaaagagagaccag1ttccatctctgtctctcaga	1676
Db	5211	ATTGATCTCAAAATGCTACTTCCCTCCCAAGAGAGCCAAATGACATCTCCGTCGTGACAG	5270
Qy	1677	caactgttcatataaacaat1tgaagcgaaggttctctgcagtagcatcatcagaaccaatg	1736
Db	5271	CACGGTGTATTAATAATAAAGCGTAAAGTATCATATCCATGATATCATACAGTACTAAATGG	5330
Qy	1737	cgtatccacgt1atatagaacagctgtctctcccaaaactgtctatcacccccaaaga	1796
Db	5331	GATTGTTCATATCATAGACAAAT1TGGTATCTCCCAAAAATTTTGCTTATCACTCCCAANGA	5390
Qy	1797	tgctctggcgaaggtctcttgcnaaatct1taactacagt1tggcagaacaacacgatalacaa	1856
Db	5391	CAACTCTGGAGAAATTTCTGCAAAATCTTACGACTTTGGCAACAAACATGGCTACATCA	5450
Qy	1857	altcaagaagt1btatatacagaactcagaagcttgcctgtcag1tcalactagctatccatccac	1916
Db	5451	ATTTAGCAACTTATATACAGACACTCAGGTTTGGTGAOTGCATCACCGATCCCATCCACAC	5510
Qy	1917	cccaagtaactgtctctcgtgcctcggaaagaacct1tggaaagctctgcccccaagaacga	1976
Db	5511	CCCAGTACTCTCTTCTGGGCCACACGACCAAGCCCTCCATGCGCTACTCTGTAACAACA	5570
Qy	1977	ggactctcgt1tcaatcaagaacaacaagaagaacagt1tgaagcttcaactc1tgaagttccagt	2036
Db	5571	GGAATTCCTGTTCAACCAAGACAAACAAGACAAGAGTGAAGAGATTTTGAAGTTTACGT	5630
Qy	2037	galtccgaagctccaagc1t1ttagctcagaactc1cccaagctcgtctctctc1tgcgaagacct	2096
Db	5631	GATACGAGATCCCAAGTTTGTACTGTGATCTTCCCATCATCCATCGCTCTGGAAGACCT	5690
Qy	2097	gcaaggtctgaagct1tgaagt1tgaaggt1tgaagct1tggcag1tgaacatccg1tgaagctcttct	2156
Db	5691	GCAAGGTTTCAGACCTGAGTGTGAATAGTGTGACTGGCAAGGCAATCGGTGTGACTCTTTCT	5750
Qy	2157	aaacgaacaat1t1tgcagat1tcaatacaccgggaactct1t1tgcgt1tgcgt1tgcgccta	2216
Db	5751	GAAATGGCAAACTGCGAGAATTTGTGAGGGGAGGAGCTCTTGTTTGAACCTGGGTGTGGCTA	5810
Qy	2217	tgcgacttgcctctca1tcaatgaactct1aacct1tga1tggcga1t1tgaactt1tactac	2276
Db	5811	CGGATGAGCTGTCTCTATTTGATCCACCTCTGGGGGGCCGCTGTGACACTTTACTAGC	5870

OY	2277	ctcgatattccgggggagagcgsgaaatgtgactttcactccaaatgcccgaacggaag	2336
Db	5871	TTTTGATGCTCTGGGGGAGTGGGAGGCTGTCAATACTCCGACGTCCCAAGCTGAG	5930
OY	2337	caagccaaagggcgtaaagaagaagtgtatctcaaacaccttaccttcaagaaagct	2396
Db	5931	TAACCAAAAGGGTGTGAGAGCAAGAGTGTCTACAAACC---TGGCTTTCAAGACGAACCT	5987
OY	2397	ggaaagcttcgcagaacacttgtcaacgttgtatccaacccccagtgctgtgcaatgtat	2456
Db	5988	GGAAGGCTGGCGGGAGAGGGTGGAGGCTGTGTATACAGATATCCGACGTCTCCGAAGGCTTA	6047
OY	2457	cttaatgtccagactgtgtcaagggccgttcggaggaccagatatcacggtttaacacgggg	2516
Db	6048	CTTGGGCGAGACTGTTCAGGCTGCGCTCTGAGAGACCAGATGCCCCGTATTAAACCGGGG	6107
OY	2517	catgtgcgcgacttcgttacaacccatgtgagacagtgtcctatgtccacacggtctcaag	2576
Db	6108	TGTGTGCTTGTGATACAGTACGTGGGCCACCGAGAGTGTAAATGTCAACACGGCTTCAATGG	6167
OY	2577	gaacgctctggaagctctgtcttgagatagggagattttggccttaactgttaagcccgacgt	2636
Db	6168	GACGGCTGTGACATGTGCTGGCGGGGAGATTGTGGCTCATTTGTGCGCTGTGGCTG	6227
OY	2637	ctccgagatgtgacaagtgtgtatagagggatcaacagcttcggggagtgccctctgtgaac	2696
Db	6228	CTCAGACCACGAGACGTGGGATGTGTGCTATCAAGGGGCTCCGGGCAGTGCTCTGTGAAC	6287
OY	2697	agggtgtgaacgacctctgtgtgtgaacatcccaacagctgtatctgcagtggtgacacctgc	2756
Db	6288	GGGTTGTGACAGGGGCCCTCGTGTGTACACTCAGGCACTTTTGGCTGTGAGTGTGTACGGCTCC	6347
OY	2757	tgtctgcgtgcaagcaccctgtgaagagaaacacacgtgtgtgtgaactgtgaactcga	2816
Db	6348	TTGTGTTGCTCATATGCCACTCTTAAGCAGACACACGCTGATGCTTAACCTGGATTATGA	6407
OY	2817	aggtgaacggatcaacatgtgacagctctgtgaactctgtccaacagaacaaaggggctgtgc	2876
Db	6408	AGGTGACGGATACATGTGCATCATCTTGTGATTTCTGCAAAACGAGCAAGGGGCTGTGC	6467
OY	2877	gaagatgtgcataagtgcttcacagaagaagcccaagtctctgtcagctgtgcaagaagctat	2936
Db	6468	AAAGGTGTGGCAGATGCTCCAGAGAGGGCAGAAAGTCTCTGTGCAAGTCCAGAAAGGATTA	6527
OY	2937	caagggggaatgcagctacagctgtgacatgaagaataaacccctgtgcagacggtgtcaacgggg	2996
Db	6528	CAAGGGGACGGGGACAGGTGCACAGATAGACCCCTGTGCAGAGGGCTTAACGGAGG	6587
OY	2997	atgcgaatgagacgcacactgtcagatatgaaggtgagggcccaaggcaagcataagtgtgaataa	3056
Db	6588	GTTGTCAAGACACGCCACTGTAAAGTGTGACAGGCCCGGGCAAGCAAAAGTGTGAGTTAA	6647
OY	3057	aagatcatatgttcggggaacggagttgaacgtgtgaagctctgaagctgtccagcgtcgtg	3116
Db	6648	AAGTCACTATATGTGGAGATATGGCTGTAACTGTATGACCGGAGACGTGCCCATATGAACGCTG	6707
OY	3117	cttacaagacaacggaatgtgccaccacagatgacagactgtgtcagaacacttactccaaga	3176
Db	6708	CTTACAGAGCAATGGCCAGTGCATGCCATACACGCCAAATGTGTGCACCTTCCAGGA	6767
OY	3177	cagacacctgagtgatatctcatctatgcgtccccaatgtggccagttacaaactgacatttga	3236
Db	6768	TATCACTGTTGGGAGTGTCCATCTATGAGCTCCCACTGTGGCCACAGTATAGCTTGTGA	6827
OY	3237	caaaagcaaaagagacctgtgtccaagaagaagctctgcagcaatagccaactacaacagctctc	3296
Db	6828	CAAAAGCCAGAGAGGCTGTGGCCAAAGAAAGTCCAGCAAAATGTGTGCACCTTCCAGGA	6887
OY	3297	ctatggccgaagaagccaagtatacaactgtgtcctcgagcggcgtgcgtctgtgaagatgtggcggt	3356
Db	6888	CTATAGCCCAAAAGCCAAAGTACCACTGTGTCTACACAGGCTGTGGCTGTGAGAGACGGGGCGGT	6947

QY	2163	acaaatgtgcaagattcaataacccgggagctcttglttgacgtgggtgtgtgcctatgcat	2222
Db	2099	CCAAACCTCGAAGAAATTCGTGACGCGGGAGCGCTCTGTTTGAACCGGGGTGTGGCCCTACGGCAT	2158
QY	2223	tgaatgtcctaactaataatccactaagctgtgcgagatgtgaacttactactcoga	2282
Db	2159	TGACTGTCTGTCTGATTGATTCGCCACTCTGGGGGGGCCGCTGTGACACCTTTACTACTTCGA	2218
QY	2283	tattccggggagttgcggaagttgcaatttccatcccaatgcccactgtgaagcaagcc	2342
Db	2219	TGCCCTGGGGGAGTGTGGAGCTGTGTCAATACTCCACAGCTGCCAAGGTGAGATAAAC	2278
QY	2403	ctgcacgaacctgtgacacggtgtgtatccaaaccccacagtgctgcagtgttaactcat	2462
Db	2336	CTGCGCGGGAGCGGTGTGACAGCTGTGTGATACAGATCCCCAGTCTCTCAAGGGCTACTTCGG	2395
QY	2463	gccacagctgtcagcgctgtcccttggaaagacagaatacacccgtgtlaacaaacggggagatgtg	2522
Db	2396	GCAGAGACTGTACAGGCCCTGCCCTGGAGACAGATGCCCCGTGTAATAACGGGGGTCGTG	2455
QY	2523	ccgcgactctgtacaacccaatggagcaagtgtcctaagtgcacacccggtctcaacgggaagc	2582
Db	2456	CCTTGATTCAGTAACTCGGCCACCGGAAGTGTAAATTCACAACCCGGCTTCAATGGGACGC	2515
QY	2583	ctgcgagctctcgtcgtgcgaatggagatttggcgctgaacttcaagcccgcaactgtccga	2642
Db	2516	GTGTGACATGTGTGCTGGCCGGGGAGATTTGGCCTGTGATTTGTTCGCTTGTGCTCTCGA	2575
QY	2643	gcattgtacagttgtatgtatggagggtatcaacagctccggggagtgctcttgtaaacaaggtg	2702
Db	2576	CCACGAGACAGTGGCATGTGATGCATCACGGGGCTCCGGGACGTGCTCTGTGAACGGGGTG	2635
QY	2703	gacacgcgctctcgtgtgtgacacactcccaagctgtatctcgagttgtgcacactgtctgtcc	2762
Db	2636	GACAGGCCCTCGTGTGTACACTCAGGCAATTTTGTGCTGTGAGTGTACGCCCTCTTGTTC	2695
QY	2763	cgtgcacgcacacctgtacaggaagaacaacgctgtgtgttaacttgaactcgaagttgta	2822
Db	2696	TGCTCATGTCCACTCTGTAAAGAGAACACAGTGTAGTTAACCTTGATTTAAGAGTGA	2755
QY	2823	cggaatcacatgtgcacagctcgtggaactctgtcacaacgaagcaacggggcctgtgcgaagt	2882
Db	2756	CGGATTCACATGCGACACTTGTGATTTTCTGCAACAGGAGCAACGGGGGCTGTGCAAAAGT	2815
QY	2883	cgctaagtgctcccaagaagaagcaccacagctctctgtcagctgtcagaagaagctacaaagg	2942
Db	2816	GGCCAGATGTCCCCAGAGAGGGACAGAAAGTCTCTTCACACTCCAGAAAGGATACAAAGG	2875
QY	2943	ggatgtgctaacgctgcgaataagaataagccocgtgtgaagcaggtgtgcacaggggagatgcca	3002
Db	2876	GGAGCGGCACAGCTGCACAACAAGATAGACCCCTGTGTGACAGGGCCCTTAACGAGAGGGTGTCA	2935
QY	3003	tgaagacgcacactgtgcaggtatgaacggccacagcaagcaataagtgtgaatgttaaaatgtca	3062
Db	2936	CGAGACACCCCACTGTAAAGTATGACAGGCCCGGCGAAGCAAGTGTGAGTGTAAAGTCA	2995
QY	3063	ctaagtccggggacggagttgagactgtgagccgttgagcagctgtccgctgcagccgtltgttaca	3122
Db	2996	CTATGTGTGAGATATGGCTGTGAATCTGTGACGGCGAGCGTGCCTCATTTGACCGTGCTTACA	3055
QY	3123	ggacaaacggaaagtgtgcaccccaatgtgcagctgtgtgcagactctacttccagggacaagac	3182
Db	3056	GGACAATATGGGACAGTCCATCAGACCCCAAAATGTGTGACCTTCACACTTCACGAGATACAC	3115
QY	3183	cgttagggatattcaatactagcgtcccccactgtggcgcaagatcaaaactgacacttgaacaaagc	3242
Db	3116	TGTTTGGGCTGTTCATTTACGCTTCCCACTGGGCCAGTATTAAGCTGTACCTTTGACAAAGC	3175
QY	3243	caaaagaaacctgtgtccaaagaagctgtgcacataagcacacttaaacacagctctctctatgc	3302

Db	3176	CAGAGAGGCGCTGGCCACAGAACCTGGACCATGGCAACCTTCAACACCACTTCTCTATGC	3235
Oy	3303	ccgaagaagccaatlatcaaccctgctcccgagccgctgctcgtgaagaatggagcgagttgaccta	3362
Db	3226	CCAGAAAGGCAAGTACCACTGTGCTCAGAGGCGTGGTGGAGACCGGGCGGTTGCTTA	3295
Oy	3333	ccgactaagatgactcctcctcaagaatgctggtgcaaaagctgttctggatcgttagactaagg	3422
Db	3296	CCCCACAGCCTTCCCTCCACAGAACTGTGGCTCTGGTGTGGGAAATGTGATGTATAG	3355
Oy	3423	atccagaagccaacaagaatgaaatgtgagatgctctctgtccggaatgaagaatgtaaa	3482
Db	3336	ACCTAGACCCACAGAGTAAATGTGGATGCTTCTCTCTATCGGATGAAAGATGTAA	3415
Oy	3483	ctgacactgcaagcgagcgtatgtggagatgctctcctgctgaatgggaacctgtgca	3542
Db	3416	CTGCACTCGCAAGGTGGGCTATGTGGGAATGGCTTCTCATGCAGTGGGAACCTGCTCA	3475
Oy	3543	ggtcccaatgctccctccctcgctcaaaactccctgacagaagtgtcgtcttccaa	3602
Db	3476	GGTCTCATGTGCTTCCCTCCACTCTCAAACTTCTGAGGGAAGTGGCTTATTCAA	3535
Oy	3603	gagctgaagcccgaggaaagcatttttgaaacccgactagccgtgcatccgtgagac	3662
Db	3536	CAGTCACTGAGTGAAGCCGTGCTATTCTAGAACCTGACTGACTGTCCATCGGGGAC	3595
Oy	3663	ccgtttgtgcgaagaagctgtggctctacacggaaataagaacccgtctcgtccggagaat	3722
Db	3596	CCTCTTGTGTGCACAGAAAGTGGGCGGGGAGATGTAGACTGTGTGGGGGGACAT	3655
Oy	3733	tgagacccaactcctaagtacagtgctcctcttcaaalgaactgtlcaatgtaacct	3782
Db	3656	CGACACCACTCCCTCCATGTGACATGTGTTCATATGACCTTGTCAATGGCCACAC	3715
Oy	3783	ctgagacatagtgtaggaagccaactgctcatctactccagccaggaccagctccacc	3841
Db	3716	CCTGCAAACGAGGTGGGAAGCAAGCTGCTCATCTGCGACGACGAGACCACTCAAC	3775
Oy	3842	--aagaagccaggtcttgatgtagaagatccatctcgtcagtggaacatcatccgcca	3899
Db	3776	GAGGAGACCAAGTGTCTTATGTGAAGAGCATTCGTGAGTGGGAATCTTTGGCTCCA	3835
Oy	3900	tgaaatccatcataatlatcttcgaaccttgtagagctctctcccaagcgacaaagcgtgc	3958
Db	3836	TGGGATCATTCATGTCTATTTCCAGACCTTTAAAGCAACCCCTGCCCCGTCACCTTGAC	3895
Oy	3960	ccaactcgtgcttgggagcaggtatatcttgtgcgtcgctccgtgtaactcgtgtgagatgc	4019
Db	3896	CCACACTGGGCTTGGAGCACAGGATCTTCTTGGCATCATCTGTGATCGGGGCTGTGC	3955
Oy	4020	ctctgacagcttaactcttaactcctcggtctaaagcagagcaactcgtgttccagcgtttga	4079
Db	3956	CTTGGCTGCTTACTCTTACTTTTGGATTAACCGGAGAAACAATCGGCTTCCAGCATTTTGA	4015
Oy	4080	-tcaagaagaagacatltgagtctctggcttlttgccaagcagcagcccaagaatatcgcaaa	4138
Db	4016	GTCCGGAAGAGCATTAATGTTCAGACTTGTGCACAGCAAGCCGTGAATAATATCTGAA	4075
Oy	4139	ccctctgtatgagaactcagcgcccggaacccccaagctcctcctgtgaaccttcaaga	4198
Db	4076	CCCCCTGTATGAGAGCAACACTCAGTCCGCCAGAACCTTCTTACGACCCCTTACGGA	4135
Oy	4199	ccctggagaagaagatctgtagaagcagccagaccctcgtggggagctgc---agctctgca	4255
Db	4136	CTCTGAAAGAGCGAGCTTGAAGGCAATGACCCCTTTGAAGAACCTGTGTGAGGGCTTGGAC	4195
Oy	4256	tgagaagccagcaagcaacaacagtcacagtttccacggtgatttccagaccagctgtct	4315
Db	4196	GGAGTGTGCACACTTACCTACTGTGCCACCTGGGCGATTCAACTGTGAATTCAGAC---	4252
Oy	4316	catgatacgaatgttctttaaagaaatgacaactcatalaagccagccatataccctctc	4375

Db	4253	CAGTTGGCCCTTTAGGACGTAAGTCTTTAAAGCACTGACGAGCATACCTCATCTCTCT	4312
Oy	4376	agttaatctggatgtgtccaggcctaaggacatgttcctcgtgatacctggggaac	4435
Db	4313	GGCTGATCTGGGGGGCTGTGTTCTGTGGTGAGAGATGTGTG-CTCTGGCCACCCAGTACA	4371
Oy	4436	tccacctcctcgtgacctat	4455
Db	4372	GCTTCTCTCTGTGACCCCTTT	4391
RESULT	3		
LOCUS	HSM801377	3642 bp	mRNA linear PRI 18-SEP-2000
DEFINITION	Homo sapiens mRNA; CDNA DKFZp434E0321 (from clone DKFZp434E0321).		
ACCESSION	AL133021		
VERSION	AL133021.1	GI:6453577	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 3642)		
AUTHORS	Blum, H., Baurasachs, S., Mewes, H.W., Gassenhuber, J. and Wiemann, S.		
TITLE	Submitted (15-NOV-1999) MTPS, Am Klopferspitz 18a, D-82152 Martinsried, GERMANY		
JOURNAL	Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de; sequenced by LMU (Ludwig Maximilians University, Munich/Germany) within the cDNA sequencing consortium of the German Genome Project.		
COMMENT	This clone (DKFZp434E0321) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de. Further information about the clone and the sequencing project is available at http://www.mips.biochem.mpg.de/proj/cdna/ .		
FEATURES			
Source	Location/Qualifiers		
gene	1..3642		
CDS	/organism="Homo sapiens" /db_xref="taxon:9606" /clone="DKFZp434E0321" /tissue_type="testis" /clone_lib="434 (synonym: htes3). Vector pSPout; host DH10B; sites NotI + SalI" /dev_stage="adult" 7..3217 /gene="DKFZp434E0321" join(<7..2607,2609..3217) /gene="DKFZp434E0321" /note="strong similarity to KIAA0246; The frame shift was determined manually" /codon_start=1 /product="hypothetical protein" /protein_id="CAB61358.1" /db_xref="GI:6453578" /translation="VGEAVGTASCKCAAGROGNTTCTAINACTISGGCSAKADCKR TTTPRRVCTCKAGTTGGIVGLEINTPGLNEHGGCDKNAEECTGTPNPAACSCILPAYTGG DGKACTLINVCLTKNGGCSFEALCNHNGVCEKNTGPNCTACBMCBGRGPDCLPC PKTSQYFEOLEHFKVDLVGPGPTVASAFADEPARVAKMDYKIMQVLHYHVV ACHOLLLENKLIENATISLOEPIVSVSSTVYINKKAKISDIIISTINGVIHIDK LSKRLNLITPKNDSGRILONTLTLANNGYIKRSNLIOGSLAVTYPDIHPVYTL WPTPOALHAIPAEQODPLFNQDNKDKIKYKLPVITPDVAVLVADLPPTSAWKLOSS ELASGCSGYNDIDPLFNGCTCRIVQELLFLDGVANGIDCLDLIDPLVGGGCTFTTF DASEGCSGVNTPSCPRMRSKFGVQKQCLINLPKRNLEGRKCSLVIOIPRCXGYV FGRQQAQPGSPDAPCNMNRGVLQYQATGECKNTGPNCTACBMCBGRGPDCLPC GCSHGQCDGIDITVDSGQCLGETGWTGSCDTQVALPAYCTPGSAHATENNTCECN LVEGDGICTCTVDSGQDNGCAGVARCSQKGTGKVCSCOKGKGDSHSTETDPCA CVDLNGGHEHATCKMTGPKGHNKCKSHYVGGDLNCPEDLPIDRCLODNGCQADAK CDLIDHPDITTVGVPHLRSPILGOYKILTPDKAREACANAAATMATYNOILSYAKAKYHC SAGLETGRAYVPTAPASQMGCSGYVTVDYGPRPKSEMMIDVFTCKMSAGAFQDL SSRKISRTIPDDLSIRGTLPVPSNGEENETLSGRILEHHLAVSWFETYNLDVNGTTP LQRFUSGILITASQDLPLETRFVYDGRALIQMDIFASNGIILHVISRPLKAPAPAT		

Query Match	Best Local Similarity	43.8%;	Score 2059, 6;	DB 9;	Length 3642;
Matches 2688:	Conservative	0;	Mismatches 729;	Indels 121;	Gaps 7;
polyA_signal	3574..3579				
polyA_site	3603				
BASE COUNT	905 a	963 c	950 g	824 t	
ORIGIN					
QY	925 aacgcgccttcgtgataccgacgcaagaagccttcgtgcaaatgtgcgacgaatccgaagg	984			
DB	1 ATCTCAGTTGGTGAAGACAGTTGGTACACCTTCATCAGTAGTGACGAGGATATCCCAAGA	60			
QY	985 aatggaacggtcgtgaacgacataatgcctcgttgagaccgaacgaatggagagtggtctaca	1044			
DB	61 AACGGGACCATCTGCACAGCAATCATCCTCTGTGAGATCAGCAATGGAGGTTGCTCTGCC	120			
QY	1045 aagcgcgactgtaaagaagacacccccaggaagaaacgggtgtgtgtgtgtgaagcgagctat	1104			
DB	121 AAGCGTGAAGTAAGAGAAACCAACCCAGAGAGGCGAGTGTGCAGCTGCAGAGACAGCTAC	180			
QY	1105 accggtgacgacatcgtgtgtccttgaaatcaaccggtgttttgagaaaccatgtgtgtgt	1164			
DB	181 ACGCGTGAATGGCATTTGTCTCCCTGGAAATCAACCCGCTTTTGGAAACATGCTGGCGTGT	240			
QY	1165 gacagaatatgcagagtgcacacagacagggcccaacacgagcggtctgttaactgtctgcg	1224			
DB	241 GACAAAGATTCGCGAATGTCACACACAGACAGGACCCACAGCGCTGTAACCTGTTTGCCA	300			
QY	1225 aagtaactgtgagatgagaaagttcgtcctgcgttatcaatgtctgcctlaacgaacaatgac	1284			
DB	301 GCATACACTGGAGATGGAAAGGTGTGCACACTCATCAATGTCGTCTTAAGTAAAAATGGC	360			
QY	1285 ggcgtgacgtccatctgtccttcgtcaacttaactgtgaagaagaatcaagatatgtacctgc	1344			
DB	361 GCGTTTACTGAAATTTTGCATCTGCATCCACCAACACTGGGCAAGTAGAAGACGCTTGACTTGGC	420			
QY	1345 aagcgacgactacaaggggtgagtggaatgtgtcgcggcgagacatcacaagggagacttccc	1404			
DB	421 AAGCCAACTACATTGGAGATGGATTTCCTTGCCTGCCGCGACAGCATTTATCAGGAGCTTCCC	480			
QY	1405 aagaacccctcgaagctcccaagtaactctctcgaagtcgaagcagatgcgtctccgaagct	1464			
DB	481 AAGAACCCGAAAAACTTCCACGATATTTTCCAGTTTCCAGGAGCAATTTGCTGAAGAATCTG	540			
QY	1465 gctggaacctgtgcctctcaaccgtgtgtcgcgccttgtctagtacctctcaatcaatgagccc	1524			
DB	541 GTCCGCCCCACAGGCCCTTTCACATGTTTTCACACTTATCTTCGACGCTTTGATGAGGAAACCT	600			
QY	1525 cggatctaaagactgtgataccagcagggcctcagttcccaaggtcttcgctatcaacgtagtg	1584			
DB	601 CGGGTTAAAGACTGGGACCAAAATACCGTTTAAATGCCCCAGGTTCTTCGGTACCACGTGTGTC	660			
QY	1585 ggcgtccagcagcgtgtgtgttgagacaacctaaagtgtgacacaagtgtgccacgaccccaa	1644			
DB	661 GCGTGCACCAAGCTCTTCTGGAAGAAACCTGAAATTGATCTCAAAATGCTACATTCCTCCCAA	720			
QY	1645 ggaagacgagtttccatctctgtctctcagaacacgtgtgtcatataaacaatgaaagcgaag	1704			
DB	721 GGAGAGCCCAAAATACATCTCCCTCTCCAGACGAGCGTGTATATAAATAAATGAGGCTTAAAG	780			
QY	1705 gtccgtctcagatgatacatatagacccaatggcgatcagccagtatagacaagtgtgtg	1764			
DB	781 ATCATATTCACGTGATATATCATCACTACTAATGAGATTGTTCTTATCATATGAGCAAAATGTGCTA	840			
QY	1765 tctcccaaaaactgtgtctatcaccccacaagaatgcctctgggcaggggtctgtcaaatct	1824			
DB	841 TCTCCCAAAAATTTTGCTTATCATCTCCCAAGCAACACTGTGGAAGAATTTCTGCAAAATCTT	900			
QY	1825 actacagtgtgcagcaaacacacgagatataccaaattcagcgaagtgtgatacaggaactcaggc	1884			

Db	911	GTGTGGAGGCTGTGTCAATATCTCCACGTGCCCCAAGTGGATTAACCAAGGGTGTGAA	970
OY	2355	gaaaagtgatcatcaacaaccggtctaaccttcaaggaggaacgttgaaggcttgcagaaact	2414
Db	971	GCAAAATGTCTCTCAACAC---TGCCCTTCAAGAGGAACCTGGAAAGGCTCGCGGAGCG	1027
OY	2415	gtgcaccgttggtgatccaaaccccaagtgtctgtccatggttaacttcatgcagactgtca	2474
Db	1028	GTGCAGGCTGTGTATACAGATCCCCAGGCTGTGCAAAGGGCTACTTTCGGCGAGACTGTCA	1087
OY	2475	gqctctgcccctggaggaccataataaccgtgttaacaaacggggacatgtgcggactgtca	2534
Db	1088	GGCTCGCCCTGGAGAGCAACATCCCGGTGTATTAACGGGGGTGTCTCCCTTGTATCAATA	1147
OY	2535	cacacccatgggaagatgactatgtccaaacacggcttcaaacggaacgctgcagactctg	2594
Db	1148	CTCGGCCACACGGAGAGGTGAATGCAACACCGGCTTCAATGGGACGGGCTGTGAGATGTG	1207
OY	2595	ctggcatgtggagatlttggccttgactgtcagcccccgacgtgtctcggagcatlgaacagt	2654
Db	1208	CTGGCCGGGGAGATTTCGGGCGCTGATTGTGTCGCCCTGTGCTCTCAGACCCAGCAAGTG	1267
OY	2655	tgtatgagggatcatcaaggtcccgggagtgctctgttgaacaaagggttgaacagcgcttc	2714
Db	1268	CGATGATATGGCATCAACGGGCTCCGGGAGTGCCTCTGTGAAGAGGGGTGACAGGCCCCTC	1327
OY	2715	gtgtgaacactcccaacagcttattctgcagtttgaacactgtgtctccgtgcagcgccac	2774
Db	1328	GTGTGCACACTTAGGCACTTTTGGCTCTGCAGTGTGTACGGCTCTTGTGTGCTCATGGCAC	1387
OY	2775	ctgtacggaagaacaacacgtgtgtgttgaacttgaactcgaaggltgaacggatcaatg	2834
Db	1388	CTGTAAAGAGACAAACACGTGTAGTGTAACTGGATTTATGAAGTGTGACGGATACATCAG	1447
OY	2835	caacgtctgtgactcttgcacaaagaacaaacggggctgttgaagggtcgtctaagtgtc	2894
Db	1448	CACAGTTGTGATTTCTGCAAAACAGGACCAACGGGGCTGTGCAAGGTGGCCAGATGCTC	1507
OY	2895	ccaagaagcaccaccaactctcttgcagcttgcagaagaagctcaaaagggtgaatgtctaacg	2954
Db	1508	CCAGAAAGGCGCAAGAAAGTCTCTCTGCAGCTGCCAAGAAAGGATTAACAAAGGGGACAG	1567
OY	2955	ctgtcatagagatlaagaccctgtgcagaacggtgtlcaaacggggatgtccatltgaacgcac	3014
Db	1568	CTGCACAGAGATTAAGCCCTGTGTGACAGGCGCTTAACGGAGGGGTGCACGACAGCGCAC	1627
OY	3015	ctgcagagatgaacggccccaggtcaagcatlaagtgtgaattgaataagtctactatgttcggga	3074
Db	1628	CTGTAAAGTGTGACAGGCCCGGCGCAACAAAGTGTGATTAAGTCACTATGTGGGAGA	1687
OY	3075	cggagttgaactgtgaagctctgaagcgttgcagcgtgcagcgttgcattacaagaacacgaca	3134
Db	1688	TGGGCTGAAGTGTAGGCCGGAGGACGCTGCCCATTTGACCCCTGCTTACAGCAATAGGGCA	1747
OY	3135	gtgcacaccagaatgcagactgttgcagaactctactctcagagacacgaacgttgaagtat	3194
Db	1748	GTGCCATGCAGAGCCAAATGTGTGCAGCTCCACCTTCCAGGATACCACTGTGGGGTGT	1807
OY	3195	ccactatcagctccccaacttgggccaagtaacaacttgaacatttgaacaaagccaagaacctg	3254
Db	1808	CCATTTAGGCTCCCCACTGGGCGCATGTAATAGCTGACCTTTGCAAAAGCCAGAGAGCGCTG	1867
OY	3255	tggcaagaagaacttgcagacatagcaaccttaacaaacagcctctctcttgcacaaagaagcaa	3314
Db	1868	TGCCAACCAACCTCTGCACATGTGCACATTAACAACCAAGCTCTCTTATGCCCCAAGAGCCCA	1927
OY	3315	gtataacccttgcctcgcgcgcgttgccttgcagaagtgtggcgggttgcctaccgaactga	3374
Db	1928	GTACCACCTGTGTCTCAGACGGCTGTGTGAGAGACGGGGGGTTCCTTACCCACAGCTT	1987
OY	3375	tgcctctcaagaagtgtgtgcacaaactgttlttggatcgtlagaactcagatcccaaggccaa	3434

Db	1988	CGCCTCCAGAACACTGTGGCTCTGGTGTGGTGGATGATGAGACTATMGACCTAGAACCCA	2047
Qy	3435	caagagtgaaatgtggaagtctctctgttaacggatgaaagaatgtgaactgcactgca	3494
Db	2048	CAAGAGTGAATCTGGGATGTCTTCTGCAATCCGATGAAAGATGTGAACCTGCACCTGCCAA	2107
Qy	3445	ggcaggctahgtgggaataggctctccgtgcaatgggaacctgcgcgaagttcctaagtc	3554
Db	2108	GGTGGGCTATGTGGGAATGAGCTTCTCATGCACTGGGAACCTGCTCAGGTCCTATGTCT	2167
Qy	3555	ctccctcgcgtccacaacttccctgacaagaagtgatcgcgtcttllccaaagagtcacgccg	3614
Db	2168	CTTCCCTCATCTCAACAACCTTCGAGCGGAAGTGTGGCTATTCACACAGCTCAGCTCG	2227
Qy	3615	aggacaagcattttgaacaacctgacagcctgtccatccgcgtggaacctgtttgtgc	3674
Db	2228	AGGCGGTCACTTCTAGAACACCCTGACTGACTGCTCCATCCCGGACCTCTTGTGTCC	2287
Qy	3675	acagaacagtggtgtacccgggaataaagaacctgtctgcggggagatltgagcacacct	3734
Db	2288	ACAGAACGTGGGTGGGGGGAATGAGACCTTGTGGGGGGGAGATCGACACCCACT	2347
Qy	3735	cactaatgtcaagcttcccttttacaatgacctgtcgaatgtgactcttgaagactat	3794
Db	2348	CGCCATGTCTAGCATGTTTTTTCTACAAATGACCTTGTCATGTGAGACCACTTCAAACAG	2407
Qy	3795	gctgggaagccaactgtcattcaacttcaagccaggaaccaactccacc--aagagaccag	3851
Db	2408	GGTGGGAGCAAGCTGCTCATCACTGCACCCAGGACCCACCTCAACCGACGAGACAG	2467
Qy	3852	gtttgtgagtgaagatgccatctctcgaggtggagatcatcgcggcaatggaaatccca	3911
Db	2468	GTTTGTGTGATGGGAAGAGCCATTCGCACTGGGACATCTTTGGCTCCAAATGGGATATTCA	2527
Qy	3912	tatttcttgaacctcttgagagcttccctccacaagcagaacagcgtctgcacctctgcgct	3971
Db	2528	TGTCACTTTCAGGCTTTTAAAGACACCCCTGGCCCCGGTAGCTTATGCCACACATGGCT	2587
Qy	3972	ggggaacagatattctgtgcgctgcgtctcgttcaactggtgcattgctctgcgcgtta	4033
Db	2588	GGGAGCAGAGGATCTTCTTTGGCATCATCTGAGTACGTGGGCGCTGTTCCTTGGCTGCTTAA	2647
Qy	4032	ctcttacttccgcttaaaagcagcgaacctggtttccagcgttttga--taagaagaga	4099
Db	2648	CTCCACTTCTTGGGATTAACCGGAGAACAACTGGCTTCCACATTTTGAATCGGAAGAGA	2707
Qy	4091	catgtatcttgcgtcttggcaagcagcagccacaagaatatcgcaaacctctgtatga	4150
Db	2708	CATTATATTGTGACGCTCTTGGCAAGCAGCAGCTGAGAAATATCGAACCCTTGTATGA	2767
Qy	4151	gacctcaagcgcgcggacccccagagtcctctctgtgaccccttcacagaccttgsgaaca	4210
Db	2768	GAGCACAACCTACAGTCCCCAGAACCTTCTACGACCCCTTACAGGACTCTGGAAGACG	2827
Qy	4211	ggaatcgaagacagcgagacctctggggagcgt---cggtctgacatgaagaacagc	4267
Db	2828	GCAGGTTAGGGCAATGACCCCTTGAGGACACTGTGAGGCCCTGGAGCGGGAATATCCAGC	2887
Qy	4268	aagcacaacagtgtaacggtttccagcagtgatcttccagcggccagctgtcctatgacagtt	4327
Db	2888	CATTCACCTACGTCCACCTGGGCACTCACTGTGATTTCTAGC--ACCAATTTGCCCTTTT	2944
Qy	4328	gttttaagaagatgaaacacctcaaaagcagcatatccatccactctgttttaactgtg	4387
Db	2945	AGGAACGTAAAGTCTTTTAAGCACTCAGAAAGCCATACCTATCTCTGTGGTGATGTGGG	3006
Qy	4388	attgtcgcacaggtcgaagagacatgttgccttgatataccttggggagcctccactcctct	4447
Db	3005	GSTGTTTCTCTGCTGGGTGAGATGATGTGTG--CTGGGCCACCCAGTACAGCTTCTCTCTCT	3063
Qy	4448	gagacctat 4455	
Db	3064	GACCCCTTT 3071	

[illegible]

OY	2168	tgccatcatcatatcacccggagccttctttagacgttgaggttgagcctatggcctatgaat	2227
Db	1231	ctctcagaaattgttcacacggggagacctctgttttgacctggagtgaggcttaaggcatttgact	1290
OY	2228	gctactcatctgaacccctacccctagcttgagccgatgtgtacaccttlaactacctcgaatlc	2287
Db	1291	gtcttcgtgatattgattccacccttgggggggccggctgttcacaccttttacttgcattgacct	1350
OY	2288	cggagggagttcggaagtttgcatcttcaatcccaatgtcccaatgaaggaacgaacaaag	2347
Db	1351	cgggggagctggggagcctgtgtcattatctccacgctcccaaggcgagctgaatcaaaccaagg	1410
OY	2348	gcgtgaagaagaatgtatctatcacccgcttaccttcttagagaggaacgtggaagcctgac	2407
Db	1411	gttgtaaacgacaaagctgtctttacaaacc---tgcccttccaagagaaaccttggaagcctgcc	1467
OY	2408	agaacctgtgacccgttgatgtatcccaaaccccccagctgtctgcatagttlaactcaatgcag	2467
Db	1468	ggggagccggttgacgctctggtgatcacatcccccagtgctgaagggcctaccttgaggcgag	1527
OY	2468	actctgaagcctgtgcccgtggaggaacgaatatacacgcgtgtataaacccgggacatgtgcgg	2527
Db	1528	actgtccag---gctgagggctgcctcttcccccttcgcacactcttaaaatgtctctcccttg	1583
OY	2528	atctgtacacacccatbtggaacgtgtgcatagtccacacccgcttcaacgggaacgcttcg	2587
Db	1584	atcagtgatctggggcaccggagagtgtaattgcacaaacggcgttcaatggagacggcgttg	1643
OY	2588	agctcgtctgcatatgtggaagatttggtgcctgtactgtcagcccgacgctgtctccgaatg	2647
Db	1644	agatgtctctggccgggagatttggtggcctgtgattgtctgcctctgtggctgcctacgacacg	1703
OY	2648	gaacagttgataagggagatcacagcgtccggggagtgacctctgtataaacagggtgaacag	2707
Db	1704	gacagtcgcgatgattggcattacggcgctccgggagtgacctctctgtgaacgggggttgacag	1763
OY	2708	ccgcttcgtgtgaacactcccaacgctgtatctgcagtggtgaacacctgtcttcctctgc	2767
Db	1764	gccctctgctgtgacactcagcagcagattttgctctgcactgtgtacgctcctcttctgtctgc	1823
OY	2768	acgcgacctgtacggaggaacaaacagctgtgtgtgtacttgaactcagaaggtgaacgga	2827
Db	1824	atggccacctgtgaaggaatacaaacgtgtgagctgaacctgtgattatgaaggctgcgaaa	1883
OY	2828	tcaatcatcacagctgtgacctctctgcacaaagacaaacaaaggggcctgtgtcgaaagctcga	2887
Db	1884	tcacatctcacagctgtgtgattttctgcacaaagacaaaggggggctgtgcgaagctgcgca	1943
OY	2888	agtgctcccaagaaagcaccccaagctctcttgcaagtgtcaagaagaagctacaaggggatg	2947
Db	1944	gatctctcccaagaaagggcaccaaaagctctctgcacgctccacgaagggaatcaaaaggggacg	2003
OY	2948	gtctacagctgcatalagatataagccccctgtgcagaagcgtgtcaacggggagatgtcacctggagc	3007
Db	2004	ggcacacagctgcacagaaatagaccctctgttcacaaacggcctttaacggaaggtgtcacgacgc	2063
OY	3008	acgcacactgcagaatgtacggggccacgaacgaatgaatgtgaatgtlaaaagtacatg	3067
Db	2064	acgcacacctgtgaagatgacacggcccgccgacaaagcacaaagtgaagttgtaaaatctcattg	2123
OY	3068	tcggagacggaatgtgaactgtgaacgtccgaagcagctgcgcgtgtgaacgttgtcttaagagca	3127
Db	2124	tcggagatgggctgtgaactgttgaaacctggagacagctgccttgaaccgcttcattacagagaca	2183
OY	3128	acggagacgtgcaccccaagatgcacgctgtgtgcagacctctactctcaaggacaacgaacgag	3187
Db	2184	atggcgacgtgtgcattgcagagcgccaattgtgtgcaccttcacattccagatatcacactgttg	2243
OY	3188	gagatattccatctagcgtcccccactgtggccagtlacaactgacatttgacaagaagcgaag	3247
Db	2244	ggggttttccactatcactctcccaactgtggccagattatgaactgtgaccttttgaacaaagcacaag	2303

OY	3248	aagctctgtgcacaaagaagctctgagacatctagccaccttaaacacactctccatctgcacaga	3307
OY	3249		
Db	2304	AGGCTGTGGCAACGAAGCTGGAGACCATGGAACCTTACACACAGCTCTCCATATGCCAGA	2363
OY	3308	aggccaagatctaacctctgtctcggcccgctctgtgagagatgaggcgagtttgcataccga	3367
Db	2364	AGGCGAAGTACCAACCTGTGCTCAGCAGGCGCTGGCTGGAGACGGGGGGGTGGCTATGCCCA	2423
OY	3368	ctacgtatgcctctcagaagctgtgtgtcaaacgcttgttggatcgttagactgaagatca	3427
Db	2424	CAGCTTTGGCTCCCGACGAAGCTGGGCTCTGTGTGGTTGGGTGATGTGACTATAGACCTA	2483
OY	3428	gggcacaaagaatgtgaaatgtgtgagatgtcttctgtttacagatgaagaatgtgaatcgca	3487
Db	2484	GACCCAAACAAGATGAAATGTGGAGATGTTCTTGCTATCGGATGAAAG-----	2531
OY	3488	ccgtcaagcagagctatgtgtgagatgtgtctctgtcagatgtggaaactgtcaggtcc	3547
Db	2532	-----	2591
OY	3548	tcatgtctctccctcgtctcacaacctctctgtacagaggtgtgtgtcttltccaaagct	3607
Db	2532	-----	2591
OY	3608	cagcccggaagacagcgcatlttttgaaacacctgtactgtactgttccatctgtgtgcacctgtc	3667
Db	2557	CAGCTTCAGGCGCGTCACTTTCATGAAACACCTGACTGACTGTGCCATYCCGGCGGACCTCT	2616
OY	3668	ttgtgcacagaagaagtgtggctacccgggaataagaagctgtgtcgtgcgcggagacattgac	3727
Db	2617	TTGTGCCCACAGAACAGTGGGCTGGGGGACAAATGAGACCTTGTCTGGGGGACATCGAGC	2676
OY	3728	accaactctaatatgtlcaacgtctcctctttaaagtacactgtgtcaatgtgaactcttctga	3787
Db	2677	ACCACCTGCGCAATGTCAGCATGTTTCTTACAAATGACCTTGTCAATGGACACACCTGCG	2736
OY	3788	ggaactatgtcgtgggaagcaactctgtctatcttaacttaagccagagaccagttcaccc---aag	3844
Db	2737	AAACGAGGCTGGGAAGCAACCTCTCATCATCTGACCGCAGGAGACCCATCTCCACCGACGG	2796
OY	3845	agaccaggttctgttgaatgtgagaatctccatctctgaatgtgagacatcatcgccgcacaaatgaa	3904
Db	2797	AGACCAAGTTTGTTGATGAGGAAGGCCATTTCTGAGTGGGACATCTTTCCTCCAAATGGGA	2856
OY	3905	tctccatlatatcttctgaaacttctgaagcttctcccaagcgagcaaaogcgtgccact	3964
Db	2857	TCATTTCAATGTCATTTCCAGGCGCTTTAAAGACACCCCTGCCCCCGTACCTTGACCCACA	2916
OY	3965	cttgcgcgtgggaagcagatatacttctgtgcgctgtctctgtgtactctgtgtcgtttctcgg	4024
Db	2917	CTGCTTTGGGAGCGGGATCTTCTTTGGCCATCATCTGTGTGACTGGGCTGTGTCTTGGG	2976
OY	4025	cagcttactcttaactctcgcgtctaaagaacgaacacactgtgttccagcgttltga-tcaag	4083
Db	2977	CTGCTTACTCTCACTTTTCGGATTAACCGGAGAACATTCGGCTTCCAGCATTTTGTAGTGG	3036
OY	4084	aagagacatltgatlctgttgcttctgttggcaagcagcagcccaagaataatcgcgaacctc	4143
Db	3037	AAGAGGACATTAATGTTGTCAGCTCTTGGCCAAAGCAGAGCGCTGAGAAATATCTGAAACCCCT	3096
OY	4144	tgatatgaaactcaagcgccgagcaacccccagagatctccctgtgaacctctcaagaacctgt	4203
Db	3097	TGTATGAGAGCAACACTCGAGCTCCGCCACGAAACCTTCTTCAGACACCCCTTCAAGGACTTG	3156
OY	4204	gagaaacagatctgtgagagacagcagccctctgggggacatg---cggtctctgaataga	4260
Db	3157	AAGAAAGGCACTTGTAGGGCAATGACCCCTTGAGGACACTGTGAGGGCTTGACGGGAGGA	3216
OY	4261	agccacaagaacaacacagatcaagcttccaaagttatccaaagccccaagctgttctcaatgg	4320
Db	3217	TGCCAGCATATCACTCTGGCACCCTGGGCGCATTAACATGTGGAATTTCTCAGC---ACCAATT	3273
OY	4321	atcagttgtttaaagaatgacaacatctaaagccagccataactcactcactcttctgtta	4380

Db	3274	GCCTTTTAGGAACGTAAAGTCTCTTAACGACACTGAGAACGCCATACCTATCTCTCGGCTG	3333
Qy	4381	atcttggagcttctgcgcagagcgtaaggacatgtctctgtatctcgttggagagcttcac	4440
Db	3334	ATCTGGGGGGTGTCTCTCTGGCTGAGAGATGTGTTG-CTTGCCCAACCCAGTACAGCTTC	3392
Qy	4441	ctcctctgaacctt 4455	
Db	3393	CTCCCTTGACCCCTTT 3407	
RESULT 6			
LOCUS	AX149459	2483 bp	DNA
DEFINITION	Sequence 15 from Patent WO0136638.		Linear
ACCESSION	AX149459		
VERSION	AX149459.1	GI:14347935	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 2483)		
AUTHORS	Shimkets, R.A., Lichenstein, H., Vernet, C. and Fernandes, E.		
JOURNAL	Polypeptides and nucleic acids encoding same		
TITLE	Patent: WO 0136638-A 15 25-MAY-2001;		
FEATURES	Curagen Corporation (US)		
source	location/Qualifiers		
	1..2483		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	619 a 648 c 639 g 577 t		
ORIGIN			
Query Match	26.5%: Score 1244.8; DB 6; Length 2483;		
Best Local Similarity	80.5%: Pred. No. 0;		
Matches 1469; Conservative	0; Mismatches 352; Indels 3; Gaps 1;		
Qy	1508	cccttaatcatbagcccgagaltaaagacttgatcagcaaggcgctcatgtccagatlc	1567
Db	571	ccctgagatcttctgtttacaggttaaaagactggagcaaaatagcgttttaatggcccgatgc	630
Qy	1568	tttcgtatcaagcttgatgtggctctccagcagctctgtcttggacaacctaaagtgcaccaa	1627
Db	631	tttcgggtaccatgtggtgcgctgcacacacactcttctggaaaacctgaattgatctca	690
Qy	1628	gtgcacagacccctccaagagagagccaggttccatctctgtctcctaagacagctgtgtca	1687
Db	691	atgctacttccctccaaagagagcccaatnagtatctccgctctctcagagacaggggtatga	750
Qy	1688	taaaacaatbagcgaaggtcctctcagtgatcatcatcagcaccaaatggcgatccacag	1747
Db	751	taaacacatnagagcttaagatcatatccagatgattatcagtgactaattggatgttcata	810
Qy	1748	ttatagacaagttgcgtctcccaaaaacttgcttatacccccacaagaatgccttggaca	1807
Db	811	tcattagacaaaattgctatctcccaaaaatttcttatctactcccaaaagacaactctgaa	870
Qy	1808	gggtctgtaaaatcttactacaaggcagcagaacaacagatataccaatccaagcaagt	1867
Db	871	gaattctgcgaaaattctttagacctttggcgaacaaacacatggcttactatcaaattttagcaact	930
Qy	1868	tgatacagagctcaggtctgcgtcgaagtcatactgaatccatccacacccagtcactg	1927
Db	931	taattacagagctcaggtttgctgagtgatcatccagatnccatccacacccagtcactg	990
Qy	1928	tcttctgacctacgcaaaagcccttgaagccttgcgcccaagcagcagaagcttctgt	1987
Db	991	tcttctgacctacgcaaaagcccttgcgcccaagcagcagaagcttctgt	1050
Qy	1988	tcaatcaagacaaagcagaagcttgaagcttactacctgaagtccaagtgcacgagact	2047

Db	1051	TCACCAACACAAACAAGGACCAACCTGATAGAGATGATTTGAAGTTTCAAGTGCATATCCGAGATG	1110
OY	2048	ccaaggtcttaagcttcagaacctcccccaaggtctgtcttccttgagaaccttgaaggtc	2107
Db	1111	CCAAGTTTAACTGCTGTGATCTTCCACATCCACTGCTGTGAAGACCTTCGAAGTTTCAG	1170
OY	2108	agctgaagtgagagtggtgaaactgtgcagtgacatcagtgagctctcttaacgaacaa	2167
Db	1171	ACCTGAGTGTGAATGTGGAAGTGGCAGGAGCATGTGATGCTCTTCTTAATGGCCAA	1230
OY	2168	tgtgcagatctataacaccgggaactcttgtttgaagtggtgtgtgacctatgcatlyaac	2227
Db	1231	CCGTGCAAGATTGTGACGGCGGACCTTGTGTTTACCTGGGTGTGGCTACGGCATTTGACT	1290
OY	2228	gctactactagaacctaccctaggtgtgcagatgtgaaccttaacttaacctgatactc	2287
Db	1291	GCTGCTGATTGATCCCACTCTGGGGGGCGCTGTGACACCTTTACTACTTTTGCATGCT	1350
OY	2288	cggggaggtgtcgggaagtgtcatcttctactcccaatgcacatgcacatgaagagccaag	2347
Db	1351	CGGGAGAGTGTGGAGACTGTTGTCATTACTTCCACGTGGCCACAGGTGGATGAACCAAGG	1410
OY	2348	gcgtgaagaagaagtgttactatacaacctgttaccttcagagaagaacgtgtgaagctgc	2407
Db	1411	GTGTGAAGGAGGAAGTCTCTCTACAAAC - --TGCCCTTTCAAGAGGAACCTCGAAGGGCTGC	1467
OY	2408	agaacctgtgcacccgtgtgtgtatcccaaccccccaaggtgtgtccatgtgttactatgcag	2467
Db	1468	GGGAGCGATGCAACCTGTGTGATACAAATCCCAAGTGTCTCAAGGCTTACTTGTGGGGAG	1527
OY	2468	actgtcaaggtctgtaccttgagaagccaagatatacacgtgttaacaacacgggtgcatltgc	2527
Db	1528	ACTGTACAGGCTGTGCTCTTGAGGACCAAGATGGCCCGTAAATACCGGGGTGTCTGCCCTG	1587
OY	2528	atctgtacacacccatgtggaacagtgctctatgccaacacggacttcaacggaacagcttgc	2587
Db	1588	ATTCAGTACTGGCCACCGGAGAGATTGAAATGCAACACCGGCTTCAATGGGACGGCGTGTG	1647
OY	2588	agctctgcctgcagatggagaatttggaccctgcacatgcagtcgaaccccgcactgtctcgaagatg	2647
Db	1648	AGATGTGCTGGCCGGGGAAATTTGGGCGCTGATTTGTCTGCCCTTGTGCTGTCTCAACCTACG	1707
OY	2648	gacagtgatgataagggagatcacaggtctccggagagtgctctcttgaacaagaggatgacag	2707
Db	1708	GACAGTGCATGATGGCATTCACAGGGGCTCCGGGCAAGTCTCTGTGAACCGGGGTGGACAG	1767
OY	2708	ccgcttctgtgtaacacttcccacaagctgtatctgcagatgtgtgacacactgtctgtccgtgc	2767
Db	1768	GCCCTTCGTGTACATCTACGCAAGTTTGTCTTCACAGTGTGTACGCCCTCTTGTCTTCCTC	1827
OY	2768	acgcgaacctgtatcgagaacaacaacgtgtgtgtgtgtacttgaacctcgaaggtgaagggga	2827
Db	1828	ATGCCACCTGTAAGGAGAACACACCTGTGTGAGTGAACCTGATTTATGAAGGTGACGGAA	1887
OY	2828	tcaacatgcacagtcgtgtgaactcttgcgaacaagaacaagggagctgtgcgaaggtccta	2887
Db	1888	TCACATGCACAGTGTGGATTTCTGCAAAOAGGCAACAGGGGGCTGTGCAAAGGTGGCCA	1947
OY	2888	agtgtctccagaagaagaccccaagctctcttgcagctgtcaagaaaggtctacaagggatg	2947
Db	1948	GATGCTCCCGAAGAGGCGACGAAGTGTCTCTCCGACGTGCCAAGAGGATACAAAGCGGACG	2007
OY	2948	gttacaagctgcataagataagacctctgtgcagaaggtgtgtcaacggggatgtgcacagac	3007
Db	2008	GGCACAGCTGCAAGGATAGACCTCTGTGCACACGGCCCTTAACGAGAGGGTGTCCACAGC	2067
OY	3008	acgcgaacctgtcagagatgtgaacgggcccagagcaagcatatgtgaaatgtaaaagtcacatg	3067
Db	2068	ACGCCACCTGTAAATGATGACAGCGCCCGGCAAGCAAAAGTGTGATGTAATAAGTCACATATG	2127
OY	3068	tctgggaacggagtgtaactgtgtgaaccttgagcaagctgtccgtctcgaacgttgtcttaacagaca	3127

	Db	2128	TCGGAGATGCGCTGAACCTGTGTGAGCCGGAGCACACTGCCATTGCACCCTGCTTTACAGACA	2187
Oy	3128	acggagcaatgaccaccaagaatgcgaactgtgagacacttaacttccaggacaagacctag	3187	
Db	2188	ATGGGCAAGTGGCATGCAAGACGGCAAAATGTGTGACTCCTCATTCCAGATATACACTGTTG	2247	
Oy	3188	gagatctccattcactgcgtcccccacttgaggccagtacaaactgtgacatttgacaagccaa	3247	
Db	2248	GGGTGTTCATCTACGCTGCCACATGGGCGCATATAGCTGCATCTTGACAAAGCAGAG	2307	
Oy	3248	aagcctgtgcccagaagaagctgtgcacataagccaactaaccaagctcttcctatgccaga	3307	
Db	2308	AGGCTGTGCCAACACAGCTGGACCAATGGCAACCTACACACAGCTCTCTATGCCAGA	2367	
Oy	3308	aggccaagatcacctcgtgctcg	3331	
Db	2368	AGACCTGGATTCCTTTTACCAAG	2391	
RESULT	7			
LOCUS	AXI49455	2512 bp	DNA	linear PAT 08-JUN-2001
DEFINITION	Sequence 11 from Patent WO0136638.			
ACCESSION	AXI49455			
VERSION	AXI49455.1 GI:14347933			
KEYWORDS	.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
TITLE	1 (bases 1 to 2512)			
JOURNAL	Shmokers,R.A., Lichenstein,H., Vernet,C. and Fernandes,F. Polypeptides and nucleic acids encoding same Patent: WO 0136638-A 11 25-MAY-2001; Curagen Corporation (US) Location/Qualifiers			
FEATURES	source 1..2512 /organism="Homo sapiens" /db_xref="taxon:9606"			
BASE COUNT	623 a 657 c 651 g 581 t			
ORIGIN				
Query Match	26.4%; Score 1243.2; DB 6; Length 2512;			
Best Local Similarity	80.5%; Pred. No. 0;			
Matches 1468; Conservative	0; Mismatches 353; Indels 3; Gaps 1;			
Oy	1508	ccttaaatcatbagccccqgattaaagaactggalcalcagcagggcctcatgtcccaagttc	1567	
Db	600	CCTGATGATCTCTGTTACAGAGTTAAGACTGGGACAAATACGGTTTAATGCCCCAGGTTTC	659	
Oy	1568	ttcgctatcacgttgtgtggtgcgcgcgcagcagctgcgtcttgagcaacctaaaagtgcaccaa	1627	
Db	660	TTTCGGTACCAGTAGTGGTGGCTGCCACGACGTCCTTG6AAAACCTGAAATTTGATCTCAA	719	
Oy	1628	gtgcacagaccctccaaggagagccagtttccatctctgctctcagagacactgtgttca	1687	
Db	720	ATGCTACTTCCCTCCAAGSAGAGGCCAATAGTCATCTCCGTTCTTAgAGCAGCGTGTAFA	779	
Oy	1688	taaaacatgatgcgaagglucctgtccagtgacatcalcatcagcaaccaatlgcgtlcatlcacg	1747	
Db	780	TAAACATAATAGGCTTAAGATCATATATCCAGTGAATPACATCAGTACTATAGGAAATTTGATTA	839	
Oy	1748	ttatatgacaagtgtcgtctcccaaaaaacttgctltacacccccaaagaatgccttgga	1807	
Db	840	TCATGAGACAAATTTATCTTCCCAAAAAATTTGGCTTATCTACATCCCAAGACAACCTTGAA	899	
Oy	1808	gggtctgcacaaacttactactacagtgtgcagcaaaacacacgatataccaaatlcagcaagt	1867	
Db	900	GAATTCTGCAAAATTTTACGACTTTTGGCAACAAACATGGCTACTCAAAATTTTACGAACT	959	
Oy	1868	tgtatcagactcaggtcgtgtcagtlcagtlcactcactgactcatccacaccccaagctacgt	1927	

Db	960	TAATACAGACTACAGTTTCTGCTACGTGTATATACACCGATCCATCCACACCCCACTACTC	1013
Qy	1928	tctctcgtgacctagcggaaacaaagccctctggaagcctctgcccccaagcagcaggactctcgt	1987
Db	1020	TCTTCTGTGCCACGACGACCAAGCCCTCCATGSCCTACTGCTGAACACAGGACTTCTGT	1079
Qy	1988	tcaatcaagacacaaaggaagaagcctgaagctcttacctgaagttccagttgactgaccgagatc	2047
Db	1080	TCAACCAAGACAAACAAAGACAACTGAAAGAGTATTGTGAAGTTTATCTGTATACGATG	1139
Qy	2048	ccaagagctttagacttaacacccctcccccaagctctgctctccggaagaagcctctgaagctcag	2107
Db	1140	CCAAAGGTTTAAAGCTGTGTGATCTTCCACATCCACTGCTCCGAGACCCCTGCAGAGTTTCAG	1199
Qy	2108	agctcgaatctgtagaggtctgtagaacctctgacgtgacatcogtgaagctctctttaaagcaacaa	2167
Db	1200	AGCTGAGTGTAATAATGTGGAGCTGGCAGGGACATCGGTGACTCTTTCTGTAATGATGCCAAA	1259
Qy	2168	tctgcagatctcaatacacccgggagcctctgtttgaagctgggctgtgctctatgagcattgact	2227
Db	1260	CCCTGCAAGATTGTGTGCACGCGGACCTCTTGTGTGACTGGGTGTGGCTTCACGGCATTTGACT	1319
Qy	2228	gcccacccaatgaactcccaactagctgagccgaatgtaacacttttbaactcttgaatc	2287
Db	1320	GTCCTGCTGATTGATTCACCCCTGGGGGGCCGCTGTGTACACTTTTCTACTTTTCATGCTCT	1379
Qy	2288	cgggggagctgcggaagttgactctttcaactcccaaatgcccacttgaagaagcaagcccaag	2347
Db	1380	CGGGGAGGTGTGGAGCTGTGTCAATATCTCCACAGCTGCCCAAGTGGAGTAAACCAAAAG	1439
Qy	2348	gcgtggaagaagaagtgtatctaaacccgctttacactcttcaagaagaagctggaagcctgc	2407
Db	1440	GTTGGAAGCAGAAATGTCTCTTACAAAC---TGCCTTTCAAAGGAGAACTGGAAAGCTGCC	1496
Qy	2408	agaacactgtgacccgtgtgtatctcaaaccccccaagtgatgcacatgatttacttcaacag	2467
Db	1497	GGGAGCGGTGCACCTTGCTGATACAACTCCCAAGGTGCTGCAAGGGCTACTTCGGGCGAG	1556
Qy	2468	actgtcagcctctgcctctggaagcagatataacocgtgtataacacccgggagcctgtgcgcg	2527
Db	1557	ACTGTCAAGCCTGCGCTGTGGAGGACCAAGTTGCCCGTGTAAATACCGGGGTGTCTGCTTG	1616
Qy	2528	atctgtatacaaccataatgggaagctgcctatgcaacacccgcttcaacgggaagcctgcg	2587
Db	1617	ATCAGTACTCGGCCACCCGAGAGCTGTAAATGCAACACCGGCTTCATGTGGACGCGCTGTG	1676
Qy	2588	agcctctgtctgacatggaagaatttgggcctgcactgttagcccccgcaagcctgtctccgagatg	2647
Db	1677	AGATGTCTGTGGCCGGGGAAGATTGGGCTGTGATTTGTGTGCCCTGTGTGCTGACACCAAG	1736
Qy	2648	gacagttgtatgaagggatatacagagctcccgggagtgctcctctgtlaaacaaggtgagacg	2707
Db	1737	GACAGTGTGATGATGGCATATCAGGGGCTCCGGGAGAGTGCCTCTGTGAACAGGGGTGGAGAG	1796
Qy	2708	cgcgtctcgtgtgacactccacacagctgtatctgcagtggtgacacactgtctgtcctgtgc	2767
Db	1797	GCCCTCGTGTGACACTCAGGAGAGTTTGTGCTGCACTGTGATGCGCTCTCTTCTCTGCTC	1856
Qy	2766	agcgcacctgtacggaaacaaacaaacgctgtgtgtgtgaacttgaacttcaagaagttgacggga	2827
Db	1857	ATGCGCACTGTGAAGGAAPACAAACGTTGTAGTGTAACTGGATTATGAAGGTGACGAA	1916
Qy	2828	tcaacatgacagctcgtgacctctctgcaacagaaacaaggggctctgtcgaaggtgcgtata	2887
Db	1917	TCACATGTACAGTTGTGTGATTCTTGTCAAAACAGSACAAACGGGGGCTGTGCMAAAGTGTGCCA	1976
Qy	2888	agttctcccaagaagaagcaccacagctctctgtcagctgtcagaagaaggtcctcaagaaggggagt	2947
Db	1977	GATGCTCCCAAGAAAGGCGACGAAGGTGTCCTGCGACGTGCGCAGAAAGGATATCAAAAGGCGAG	2036
Qy	2948	gctacagactgtacatagatatagaccctctgtgcagacgggtgtcaacggggagtgtcactagc	3007
Db	2037	GGCAGACGTTGCACAGATATGACCCCTGTGTGACAGCGCTTATTAACGAGAGGTGTCTACGAGC	2096

QY	3008	acgccacctgcaggaatgacgagcccaagcaagaataagtgtgaatgtaaaagttcaactatg	3067
Db	2097	ACGCCACCTGTMAAGTATGACAGCCCGGGGCAACACAAAGTGTAGTGAAMAATCACTATAG	2156
QY	3068	tccgggaagcagtgatgacgtgtgagcccttgagcagctgcgcgtgcagcgtgtgtctaaagaca	3127
Db	2157	TCGGAGATGGGCTGTAAGCTGTAGCCGGAGACACTCCATTGACCGCTGCTTACAGGACA	2216
QY	3128	acgagcaatgtccaccacagatgtcagaactgtgcagacacttacttccaggagacagacgttag	3187
Db	2217	ATGGGCAGTGGCATGTACAGAGCCCAATATGTGTGACTTCACCTTCCAGAGTACCACCTGTTG	2276
QY	3188	gagtatccatctacgtctcccaactgtggccagttacaactgtacatttgacaaagccaag	3247
Db	2277	GGGTGTCCATGTACGCTCCACCTGAGGGGCAATAGCTGACCTTTGACAAAGCCACAG	2336
QY	3248	aagctgtgtgcacaaagaagctgtgaccataagccacttaaacagctctcctaagccaga	3307
Db	2337	AGGCTGTGGCCAAAGCAAGCTGGACCAATGGCAACTACCAACCAAGCTCTCTATGGCCAGA	2396
QY	3308	aagcaagatcatcaactgtgctcg	3331
Db	2397	AGACCTGGATTCCTTTACCAAGG	2420
RESULT	8		
LOCUS	AXI49447	2011 bp	DNA
DEFINITION	Sequence 3 from Patent WO0136638.		Linear
ACCESSION	AXI49447		
VERSION	AXI49447.1	GI:14347929	
KEYWORDS	human.		
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
REFERENCE	1 (bases 1 to 2011)		
AUTHORS	Shimketers, R.A., Lichenstein, H., Vernet, C., and Fernandes, E.		
TITLE	Polypeptides and nucleic acids encoding same		
JOURNAL	Patent: WO 0136638-A 3 25-MAY-2001;		
	Curagen Corporation (US)		
FEATURES	Location/Qualifiers		
source	1..2011		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	458 a 551 c 583 g 419 t		
ORIGIN			
Query Match	25.4%; Score 1196; DB 6; Length 2011;		
Best Local Similarity	77.6%; Pred. No. 0;		
Matches 1524; Conservative	0; Mismatches 330; Indels 110; Gaps		
QY	1868	tgatacaggaactcaggtctgtctgcatgaatcaactgaactcaatcacacccagtcactg	1927
Db	158	TCGTTTCAGAGACTCAAGGTTGTGTGAGTGTATCATCAGCATCCATCCACACACCCAGTCACTC	217
QY	1928	tcttctgtgactacgacaaagcccttggaaagctctgcccccaagacagagactctcgt	1987
Db	218	TCGTTTCGCGCCACCGACCAAGCCCTTCATCTGCTTCACTGCTTAACACAAACGAGCTTCGT	277
QY	1988	tcaatcaagacacaaaggaagctgtgaatctctactgtgaagttccacgtgtacacagact	2047
Db	278	TCACCAACACACACAGGACGACGATGTAAGAGATATTGAGATTTCATGTGATACGAGATG	337
QY	2048	ccaaagcttgaactcagaactcccaagctcgcctctctctcgtggaagacccctgcagactcag	2107
Db	338	CCAAGGTTTACTCTGATGATCTTCCACATTCACATGCGCTGGAAAGACCTCGCAAGGTTGAG	397
QY	2108	agctgtgtgtgtgaggtgtgtgaactgtgcagtgatcatcgtgtgactcttctctaaagcaaa	2167
Db	398	ACCTGAGTGTGAAGTGTGAGCTGGCAGGAGCAATGGGTGACCTCTTCTGATGTGGCCAAA	457

OY	2168	tgagcagatcacaacaccgggaacctglttgaacgtgggtgagccctatgacatgact	2227
Db	458	ccctcacaattgttgacacgggagacctctgtttgacctgggttgacctgaacgcatgtact	517
OY	2228	gcctaccatagaatcctacccttagtgagccgatgtagacacttactaacttgatatac	2287
Db	518	gtctgcgatattgaattccacaccttggggggccgctgtgacaccttttactcattgcctgct	577
OY	2288	cgaggagatgcggaagcttgatcatttcaaccccaatggccacttgaagagcaagccaag	2347
Db	578	cggggagactggggacctgtgtcaattactccacagctgcccaagctgagatgaaccaaagg	637
OY	2348	gcgtgaagaaagatgfatatacaaccgctlaaccttcaagagagaacgtggaagctctcc	2407
Db	638	gttggaagcagaagatgctcttaccaccc---tgcccttaagagaaacctggaaagcttccc	694
OY	2408	agaaactgtgcaacgtgtgtatccaaaccccccaagtlgtctgcaagtattcatalgccc	2467
Db	695	gggagccatgcacagctgtgctatcacaattccacagctgtcgaagggctacttccggcgag	754
OY	2468	acttcaagagctgcgcctggggggagacagatagacccgtgtgaacaacgggggcctlggccc	2527
Db	755	actgtcagagcctgaccttgaggagaccagatgcccccggtatattaccggggctgtcttcc	814
OY	2528	atctgtacacacccatlggagcaatgtgactatgtcacacggcttcaacggagcaagcttcg	2587
Db	815	atcagtatctggggcaccggagactgtatattgcaacacccgcttcaattggagacggctgtcg	874
OY	2588	agctcgtcgtgacatgggaatlttgagctgtactgtcaagccccgaagctgtccgaagatg	2647
Db	875	agatgtcctctggccggggagattttgggcccgtgattgtctgcctctgtggctctacagccag	934
OY	2648	gaagagtgtgtaggggagatccagagctccgggggaagtgctcctgtgaacaaggtgtgagca	2707
Db	935	gacagtcgcatgatggcatatccgggctccgggagtgctctgtgtaaacgggggtgagcag	994
OY	2708	ccgacttcgttgacaactcccaagcgtatctcgcaatgtgcaacactgtctgtccgtgc	2767
Db	995	gccctctgctgtagacactcagagcagttttgtgtcgcaatgtgtacgctctcttcttctgcctc	1054
OY	2768	acgcacacctgatacggagaacaacagltgtgtgtaacttgaactagaagtgtagcyga	2827
Db	1055	atggcacctctgaagagagaaacacacgtgtgagtgtaacttgattatgaacggtgacgaa	1114
OY	2828	tcaatgacacagtcgtgtgacttcttgaacaacagaaacaaaggggctgtgcgaagtgccta	2887
Db	1115	tcacatgtcacagattgttgattttgtgaaacacagacaaacggggcctgtccaaaggctggcca	1174
OY	2888	agtgtctccagaagaagcaccgaactctcttgcagctcgtgaagaagaagctcacaagggagatg	2947
Db	1175	gatctctcccaamaggcacaagaagttctctgcacgctgccagaaaggatatacaaaagggacg	1234
OY	2948	gctacagctgcatalagatagatagaccccttgtgcagaacggttcaacggggagatgacatgac	3007
Db	1235	ggcacacagctgcacagagatagacccctctgtgacagagggcctttaaggagaggtgtcacagac	1294
OY	3008	acgcacacctgacgatagacgggcccacagcaacataagltgtgaaatgtaaaagtaactatg	3067
Db	1295	acggcacctctgaatgatgcacagcccccggcagacacaaagtgtagtgaatgtaaaagtctactg	1354
OY	3068	tcgaggaacgagatgtgactgtgaagcctggaacagctgcgctgtaacgctgtgtactaagaca	3127
Db	1355	tcggagatggcctgtaactgtgagaccggaagacgttgccatttatgacgcgtgctttacagagca	1414
OY	3128	acggaacagtgcaaccccaagcagctgtgtgcagacacttacttccaggaacagacgtgag	3187
Db	1415	atgggacagtgacatgcagagacgccaaattgtgccgacctccactttccagatgaatgacactgttg	1474
OY	3188	gagtatctccatctaacgctctcccaactgtggccagtlacaacatgacattgtgacaagccaag	3247
Db	1475	gggtgtttccactttacagcttccctccactggggcagatgaatgaactgtgaccttttgacaaagcagac	1534

QY	3248	aagcctgtgtccaaagaagctctgcgaacatagccacactcaaacagctctcctactgtcccgaga	3307
Db	1535	AGGCGCTGTGCCAACGAAAGCTGTGGACCATGTGGCAACTACAAACAGCTTCTCTATGGCCAGA	1594
QY	3308	aaggccaagatataccctcgtctcgtccgcgcgtctgtcgttggaagctgtggcggggtctgcctaccga	3367
Db	1595	AGGCCAAGTACCAACTCTGTCTCAGCAGGCTGGCTGTGAGACCGGGCGGGTGTGCTTACCCCA	1654
QY	3368	ctacgtatctcctcctcagaagctgtggtgcgaacgcttctgttggatctglaagactcggatcca	3427
Db	1655	CAGCCTTTCGCTCCAGAACCTGTGGGCTGTGGTGTGGTATGATGAGCATGTGAGCCCA	1714
QY	3428	gggccaacaagaagtgaatctgtggatctgtctctgttaccggaatgaagaatgtgaacttga	3487
Db	1715	GACCCAAACAGAGTGTAAATGTGGATGTCTTCTGTGATGCGATGGAAG-----	1762
QY	3488	ccttgaaggcagactatgtggagatctgtctctgtgaagctgtggaacctgtcgaagctcc	3547
Db	1763	-----	1762
QY	3548	tcatgtctctccctcgtctcaacacttctcagagaagctgtcgttcttccaaagact	3607
Db	1763	-----GAACTGCTGGCGCTATTCCACAGCT	1787
QY	3608	cagcccgagagacagcatttcttgaacacacctgtacctgtctccatcgtgtgacacctgt	3667
Db	1788	CAGCTCGAAGCGCGTGCATTTCTTAGAACAACCTACATGACCTGTCCATCCCGGACCCCTCT	1847
QY	3668	ctgtgcacagaacagctggcggtacccgggaataaagacctgtctcgcgcggagacattggac	3727
Db	1848	TTGTGCGCACAGAACACTGTGGCTGGGGGGAATGAGACCTTGTCTGTGGCGGGAATCGAAC	1907
QY	3728	accacctcaactaatgttcaacgctctccttcttacaatgaacctgttcaatgttacccttctga	3787
Db	1908	ACCACCTCGCCAAATGTCACACATGTTTCTTCTACAAATGACCTGTCAATGAGCACACCCTTGC	1967
QY	3788	ggactatcgtctgtgaagcaacatgctccatctactccttaagccaggaac	3831
Db	1968	AAACGAGCTGGGGAAGCAAGCTGCTATCATCTGCCAGCCAGCGAGC	2011
RESULT 9			
AX149449	1804 bp	DNA	linear
LOCUS	AX149449	1804 bp	DNA
DEFINITION	Sequence 5 from Patent WO0136638.		
ACCESSION	AX149449		
VERSION	AX149449.1	GI:14347930	
KEYWORDS			
SOURCE	human.		
ORGANISM	Homo sapiens		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
REFERENCE	1 (bases 1 to 1804)		
AUTHORS	Shmukler, R.A., Lichenshtein, H., Vernet, C. and Fernandes, E.		
TITLE	Polypeptides and nucleic acid encoding same		
JOURNAL	Patent: WO 0136638-A 5 25-May-2001;		
FEATURES	Curagen Corporation (US)		
source	Location/Qualifiers		
	1..1804		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
BASE COUNT	426 a 487 c 521 g 370 t		
ORIGIN			
Query Match	21.6%; Score 1016.4; DB 6; Length 1804;		
Best Local Similarity	82.0%; Pred. No. 2e-279;		
Matches 1183; Conservative	0; Mismatches 256; Indels 3; Gaps 1;		
QY	1868	tgatacagacccagcgtctgtctgtcaatcatcaatgaactcatcacacccagcagtctg	1927
Db	158	TGTTTTCAGAGCTACAGTTTGTGAGTGTACAGATCCAGATCCATCCACACCCAGATGATC	217

[illegible]

D	4885	CAAGGAGGATAGTGGAGTGGGACAGTGTGTGTGCGCTTGAAGGTTTGGCTGGGACAGC	4944
O	2583	ctcgagactctgtctgscataygaagatttgggcctgaactcaagcccgagctgtctcga	2642
D	4945	CTGTGAATCTGTGTCCTCTGTCCTGTCCTTTGGGCCCCCAATTGTCAAGCCCTGGCTGCACGT	5004
O	2643	gcatgagcaagtgtgatgaagggagatacaagactccgggagatgaactctgtgaacaagagt	2702
D	5005	CGATGTGCCCGCTGTATATGAGGCCCTTGGGGGCTCTGGCTCTCTGCTTCTGTGATGAAAGCGCTG	5066
O	2703	gaacgcgactctgtgtgatacaactcccaagctgtatctgcagttgtgaactgtctgtc	2762
D	5065	GACTGGGCGCACGCTGTATAGGTTGACACTGGAGGCTGCAGCCCTGTGTATACCCCACTGTGC	5124
O	2763	cgtgacgccaacccgtatacggagaacaagaagtgatgttaactgtgaactaagaagtga	2822
D	5125	ACCCGAGGCTGTGTGCCGTGACGAGCAACACTGTGATGTCAGCTTACCGCTAATGAAGGGGA	5184
O	2823	cgagatcacatgacaagtcgtgtgactctgtcaacaagaacaacgggggctgtgtgaagt	2882
D	5185	TGGCCGCTGTGTGTACGTGACAGCCCTGTCCAGGACGGGGCATGGGCTGCATGACGA	5244
O	2883	cgctaaagtcctccagaagaagccccaagctcttcagcttcaagaagaaggtataaagg	2942
D	5245	CGCAACTGTAGCCCATAGAGTAATGTGATCTTACTGTACTGCTGCCGACTTACGAGGG	5304
O	2943	ggaatgctacagctgtcgaatagatagacccctgtgcaacggtgttcaaacggggagatgca	3002
D	5305	TGATGTGCTGGAGCTGCCGGGCGGCAACCCCTTGACAGATGGGCACACGGGCGGCTGCAG	5364
O	3003	tgaagacgcacactctcagaagatgacgggccagggcaagcataagtgtaatgttaaaagca	3062
D	5365	CGAGACACCCCAACGTGTGAGACACCGGCTGAAACACACGGCGCTGTGACTGTCCACGACGG	5424
O	3063	ctatgtcgggaagcgaatgagactgt---gagcctgagcagctgtgcctgtgaactgtgt	3119
D	5425	CTACGTTAGCGATGTGACTGTGCAAGTCTGTGGAGAGTGTGGAAACACCTGTGTGACCGTGTCT	5488
O	3120	acagagcaacgagacagtgtccacccaagatgtccagcgtgtgcagacctctactcagaagac	3179
D	5485	GGGCGAGCACCGCCCTGGCCCACTGATGTGCATAGTGCATGCACTGTCACTTCACATTTCCAGAGAA	5544
O	3180	gacgtaaggaatattcatatctacgctcccccacgtggcgcaagtaaacatgatacttgaca	3233
D	5545	ACGGGCTGGCTTTTCCACTCTCCAGGCCACACAGGGCCCTTAATGGCTAATCTTTTCGA	5604
O	3240	agccaagaagcctgtgtccaaagaagctgtgcacatagccaactacaacacgctctccta	3299
D	5605	GCGTACGCGCATGCGCAAGACACGGAGCGCTCTCTTCACTTCCCTCACCTCTCTGCG	5664
O	3300	tggccaagaagccaagatacaacctgtgtctgcgcgcgtgtgtcttgagaagtgtggcgttgc	3354
D	5665	TGCCAAGAGAGCTGGGCTTCCACTGTGCTCATAGGCTGTGGCTGTGGCAATGGCTCCACTGC	5724
O	3360	ctaacgcgaactcgtatgcctctcagaagtgtgtgcaaacgctgttggagatcgttagacta	3419
D	5725	CCACCTGTGTGTTTTCCTGTGGCGGACGTGTGGCAATGTGCGGGGTGGCATATGTACGCT	5784
O	3420	cggatccagggcaacaagaagtgaatatgtgtgagatgtctctgttaccggataaagaagt	3479
D	5785	GGGTGCCCGCAAGAAACCTCTCAGAACCTGTGGATGGCTTACTGTCTTCGTTCCAGATGT	5844
O	3480	gaactgcacccgcaagcagactatgtgtggagaatgt---cttctgtgaagtggagaact	3536
D	5845	GGCTTCGCGATGGCGAAATGGCTTTCGTGGGTGACGGGATTCAGCACCTGCATGGGAAGCT	5904
O	3537	gtgtcaagtctctcatgtctctccctctgcacaaactcttcagacaaggtgtgtgctt	3596
D	5905	GCTGTGATTTGCTGGCTGGCACATGCGCACTTCTCCACCTTCTAATGGATGCTATTGGGGTA	5966
O	3597	ttcaagaagctcagcccgagagacagcattttgaaacacactgtacgtacactctgtcatcgg	3655
D	5965	TGCCAATGCCACCCAGCGGGGTCTCGAATCTTCGAGACTTCTGTGATGAAGAGACTCAGTA	6024

OY	3657	tggacccttgcttgccacagaacagtgagcctgaggaataagagccgtgctgagcg	3716
Db	6025	TAAGACACTCTTGCCTCCCTGTCAATGAAGCCTTTGTGGACACATGACGCTGAGTGGCC	6084
OY	3717	ggacattgagcaccactactaataatgcaagctcctcttactaagaagc-----cct	3768
Db	6085	AGACTGTGAGCTGACATGCTCCCAAGCCACCTCTTAAGTGCCACACCCACGACGAGGAA	6144
OY	3769	gtcaatgtaaccttcttggaagactatgctgaggaagccaactgctactactcagcag	3828
Db	6145	GTTTGCTTCCGGCCACTCAGGCTTCAACCTCATCATCATGACGACGAGCCCTGACAAAG	6204
OY	3829	gaccagctccacccaagaagaagagctt-tgtgataggagaatccactcagtcagtgagac	3887
Db	6205	TTCCTGGGCCCCCTGTGGCCCCAGGAGAGTGTGTAGCCGTATCATGTGTGGGACAT	6264
OY	3888	catcgccgccaatgataatccatatttctgaaccttgaagagctccctccacgac	3947
Db	6265	CATGGCCCTTCAATGGCATCATCATCTGTGGCAGCCCTCTGTGACACCCACAGCC	6324
OY	3948	agcaagagctgcccactctgagcctgaggaaggtatattctgtgcgctgctgtcac	4007
Db	6325	CGCAGTGTGCGCCCTTGAAAGCCCCACCTGTGGCGGAGGCGTGGGCTGTGCTGCCGC	6384
OY	4008	tgtgtgc 4013	
Db	6385	TGGAGC 6390	
RESULT 11			
LOCUS	HS275213	7870 bp	mrna linear PRI 29-MAR-2000
DEFINITION	Human sapiens mRNA for stabilin-1 (stabl gene).		
ACCESSION	AJ275213		
VERSION	AJ275213.1	GI:6469373	
KEYWORDS	stabl gene; stabilin-1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
AUTHORS	1 (bases 1 to 7870)		
TITLE	Pollitz,O., Guillot,P., Gratchev,A., Schledzewska,K., Birk,R., Hakly,N., Tebbe,B., Orfanos,C.E. and Goerdts,S.		
JOURNAL	stabilin-1: an endothelial-macrophage member of the fasciclin domain containing protein family associated with angiogenesis		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 7870)		
TITLE	Pollitz,O.		
JOURNAL	Direct Submission		
REFERENCE	Submitted (25-NOV-1999) Pollitz O., Department of Dermatology, Benjamin Franklin Medical Centre Free University Berlin, Hindenburgdamm 30, Berlin 12200, FRG		
COMMENT	Related sequence D87433.		
FEATURES	Location/Qualifiers		
source	1..7870		
gene	/organism="Homo sapiens"		
CDS	/db_xref="taxon:9606"		
	31..7743		
	/gene="stabl"		
	31..7743		
	/gene="stabl"		
	/codon_start=1		
	/product="stabilin-1"		
	/protein_id="CAB61827.1"		
	/db_xref="GI:6469374"		
	/translation="MAGPRGLPLCLFLAFLGAFSEFVROVLFGKGDVKTFTVTHVPC		
	TSCAIAKQOTPCSGMLRELDPDITQDCREVLQGSMSMSCRRRCRAQVOKACP		
	GYWGRCHCEPGGAETPCNGHGTCLGDMRNGTCQENFRKSGACEODPNRPFDC		
	OSVCSVHCVENHGRGDGSCICFAGYTGPHGDOELPVCELRCPQNTCCSEAPSCR		
	CLPGYTOGSECRAPNCPMPCSPCLLAOCVSYPKQAOCHCEENHGDGMDCLPDC		
	TDNLGGCPSNSTLYQYKRGQAFCTCRPOLVINSNASAGCFAPCSFSCDSATCOV		
	TADKTSVCVRESEVGGKACYGHLHLEVOYKATGTGRVFLQIRVAVAMMDGCRELT		

OY	25	ctgagacagatgcccgactatccatttccgaggttactatlaactaacaactgagca	84
Db	3433	CTGGACTTGGTGGCTCCCTCAGCCCTCTCCGGAAATTGCTGACGACCATGGGTGTG	3492
OY	85	atgcaatcgagctcgagctgactatcgtgtgtgccaacaatgaaagcatcgaa	144
Db	3493	CCCCAGATTGAGGCTGCACCTGACCATCTTGTGCGCCACCAACCCCTCCCTGAG	3552
OY	145	aactatatacagagagaagaagccacatctctaaagaagataattcagtgatacattg	204
Db	3553	GCC-----CAGGCAACAGCAGTCACTGAGCCACACACAGTGGCACCATGTG	3603
OY	205	gtccctgaggaagaagctcctgaagaatgacttgcataagggatgcaacgagagaccatg	264
Db	3604	GTCCCTGGGGAAGGCCCTCTCCATGAAACCTCGGGAAGGTGGACCCCACTCCCTC	3663
OY	265	ctggaggtctccctaccctctgctcttcttccgcaatgagcagctgtatgtaaaagaa	324
Db	3664	CTGGCCCTCCCACTGGATGCTTCTTACACACACAGTGGCCAGCCTGAGTGAACAT	3723
OY	325	gtctcaataaactacaccaaigtgccaactgataaagagtgatccaatgtctgagga	384
Db	3724	GTGGCACTGGAAGGCCCATAGTGTGAGAGGCCCTGGGCGGTGGATGTGTGTGGGG	3783
OY	385	gtctctgaaatctcaagaagaagatgtgacataatgacaacattatgtgagagggag	444
Db	3784	GTCTTGACGCTGGCTCAAGTCCCTGCTTCAATGAGCCACGCTGAGCCCTGGCGGAGAA	3843

BASE COUNT 1420 a 2482 c 2446 g 1522 t

ORIGIN

Query Match 14.8%; Score 697.2; DB 9; Length 7870;

Best Local Similarity 51.5%; Pred. No. 6,9e-188;

Matches 2066; Conservative 0; Mismatches 1823; Indels 125; Gaps 15;

1AGPPTVLVPSVSSFSRTMNSLAOOLCRHIIAGOHLEDTRTQOTRRMTLACOE
ITVTFMQFTKYSYKYKDQPOQFNIRKANNIANGVPHVYTGILRMOAPSGTPDPKRT
IGQILASTFAESREFTILRNCGLPSLIDGGEPTVPAFVPAEVAEDVLRJDLATLPLFAG
LSKLOELVRYHIYLNHGLTVERKILSPKRIKTLTMAOVLAVNISEEGTILICEPVRIOR
VDYMANGVYIHMLDGILPPTLPIPIPKHSEORHIVAGSCYDCAALNTSTPPNSV
KIDIFPEKCVYIHDPTGLVNLAKKGCASTYCNQOTMEGCKGFFGPCTCCPGFSNFC
YKGNCSDEIOGNACILCPDYKGIACHICISNPNKNGEDOCEDGCVHGLCDNRPSG
GYOQOCTGACFSGFRCNESMGDCGPTGCLAQHNLHARCVSGGVARCRDLDFEEDG
FSCPTNPSCHPDGGCSSENAECVPSLDGTTHCTCKHGMSDRGVRLADECLDVG
GCTHDALCSYVPRGRCCKLGEFAVAGVOCSPIDPCRMNGGCHCLATCRVAGGOR
VCTCPRGFGDGSCTGDIJRELLEAANHSTFYOMLKSAITLPIADRYATLVPSEA
VROLSPEDRAFWLQPTPLPNUVRANHLQVALFEELARLGGDEVALDLPTEWELRNI
SGRWYONASVYADLADLNTNGVLIHLISLOALPLPRGDVPGGGLLOLDLDPVAFSLRE
LLOHGLVPOIEAATATYTFVPPNRSLAEQNSHLDADTRHRHVVGLALSMETLRK
GGHRNSLLGPAMHIVFVNHSGDEPVNHVPLBEGMLPAPRSLIGLSGLVLYVSSRLH
SHAEALREKVCNCTRRFCTGTGLODTPRKSVCYVRSFSGSFSGCTYCTAKKIQYVDC
CFEFTGLCEPFGGAGVSGSGHGCCODNPLSGSECHGHEGHTGCEVELGRIAPN
CTGVCDCAHGLCOEGLQGGSCVGNWMOGLRCDOKITSPQCRKCDPANVCYODSAG
ASTYCAAGYSGMNGIFPCEVDPDCAHGGCSPHANCCTKYAPGQRTCTODGYMGDEL
COEINCLLHHGGGCIHAREIPTGPOVYSCSCEGSGDGIPTCELLDPCSKMNGCS
PYATKSTGSDGORTCTCDTAHYHMGDLTGRARYGELLEDRKHAFFSLRLIEYKELKG
DGPFTTFVPHADLSMSODELARIRAHNQLVRYHYVRCRLRSEDLIEQYATRLS
GHPLPSEDEGSTLYLNDFAVVSSDHEAVNLHFTDRVLLPEALHMBEDAPLPIR
NMTAAAGFGYKIFSGLLVAGLLPLREASHPPTMPTDAFPAALPDRQAMLYH
EDHRDLKLAILRQHMTIRNVEALASDPLNGLPRLTMGTPISECSRTRGELMVGEDD
ARLVORHLPFEGGLAYGIDOLLEPGLIGARVWHPSLIMGROGLRGCRVNTTMMKPS
CCPGHGSCEQACRGPPSSPCSDRBYCMOMSGSCCCLRSQFAGTACELCAPAGAP
HCQACRTYHGRDEGLGSGSCFDEGWTGPRCEVOLELDQVCTPPCAPEAVCRAN
SCESCGLYEGDGVCTVADLCQHGCGSEHANCISOVMTCCTPLPDEGDMSCRA
RNPCTDGHRRGCSSEHANCISLGTINTRECEHAGYVDGCLLEESPPVDRCIAGOPP
CHSDAMCTDOHFOEKRAGVFLQATGSPGLNFESEAEAGCAVLAPEPOLASAOO
LGFHCLLMGMIANGSTAHVYVPEVADCGGSTRVYSLARKNISEWMDVCFYFDOVA
CQRNFPEVVDGISTCNKGLLDVLAATRANSTTYGMLLSTGANRORLDPDLDEIDLT
YKTLFVYVNBGEVDNMTLSGPDLEHLASNAATLLASNASOGKLLPAHSGLSLITSDGP
DNMSWAPVAPGIVAVVSRIYVMDIMAFNGIITHALASPLAPQOVALAEAPVAVAG
GAVLAAGALIGVAGALYLRANKRPGFPGFSAFOADDADDPSFMOEGTNPLTVSVP
NPVFGSDTFCPEPDDSLLEEDPDTORILTVK"

OY	445	tgtggaagatggttcccaagcaagcccttcgcccactcgcgagacaaaacacattagaagagc	504
Db	3844	TGTTTAATACGTGCAACCAAGAGATTTCCGCTCACTCAAGGGCTTCCAGCTGCAGACACACC	3903
OY	505	aggaatgcatactatccatctactcaatcggggaagagatccgtaattcatcggtgcgcag	564
Db	3904	AGGAAGAAGCTGTGTCTACCGATCTGGCTT-----CTCTCTTCCCGGGCTGTCT	3954
OY	565	ccacagatgtgtggaacatcatctacaagagccctgtcgtctgtctctttggccaca	623
Db	3955	TACACATGTGCCCAAGAAAGATCCAGGTGCGGAGCTGCTCCCTGTGGTTCTTGGACACCTG	4014
OY	624	tgccaagccctgcgcccgggaaggtcaaaaatgtgtctctgggaacggtctctgtcgtgga	683
Db	4015	TGTAGCGCATGCCCCAGGGGCTCTTAGGGGGGGTGTGCTCAAGGCCATGGGCACTGCC-AGGA	4073
OY	684	cgtgtgaatgtgcactgtgcacgtgtgcagttgcggtcgttcaattggaacgctgtga	743
Db	4074	CAGTTTCTTGGGACGCGGGAGTGGTCCACATGCCACAGAGGGCTTCCATGGAMAGGGCTGTGA	4133
OY	744	aactgtcaactggagggaagtatgtctccactgcgcgaaccaagatgctcttgtccatg	803
Db	4134	GCTGTGTGAGCTGGGCGCGCTACGGGCCCAACTGCACCCGAGTGTGACTGTGCCCCATGG	4193
OY	804	gagaatgacaagaagaccttggagaacgctctcgttgaactgtgaagtcgcgttcggcga	863
Db	4194	GCTGTGCCAGGAGGGGCTGCAAGGGAGCGGAAGTGTGTCTTAACGTTGGGCTGGCAGGG	4253
OY	864	agtgaagtgtgacatgtgagatgccacacagaaacatcgtcaacggtccttcaacacagatgc	923
Db	4254	CTCTCCGCTGTGACCAAGAAATTCACAGCCCTCAGTGCTCCATGGAATGTCAGACCCCAATGC	4313
OY	924	caactgcctctctgatalccaagcgcgaagccctcgtgcgaatgtgcgcaagatccga	983
Db	4314	CAACTGGGTGACGAGACTCGGCGGAGCCGTCACATCGCGCTGTGCTGGGGATTCCTCGG	4373
OY	984	gaatggaacggtctctcacagccataatcctcgttgaagccagcaattggaagtctctac	1043
Db	4374	CAATGGGATCTTCTGTTCACAGGTGAGCCCTGCGGCCACGGGCCATGGGGGCTCTCCCC	4433
OY	1044	aaagccgcactgttaaaaagaaacccccagaaacgggtgtgtgtgtcaaggaagcta	1103
Db	4434	TCTATGCCAATCTATACCAAGGTGGCACTCTGGCAGCGGAGCATCACTCCACAGATGGCTA	4493
OY	1104	taccgcgcagcggcatcgtgtccttgaataccaacccgtgtttggaacaatggtgtcgt	1163
Db	4494	CATGGGCGAGGGGAGCTGTGTGCCAGGAATTTAAAGCTGTCTCATCCACACGGGGGCTG	4553
OY	1164	tgaagaatgtcgagatgtgcacacagaacgaaggcccaacgaagcgctcttaactcttgc	1223
Db	4554	CCCACTTACGCCCGAGTGCAATCCCCACTGACCCTGCCCCAGCAGGCTCTCTCACTGCGCTGA	4613
OY	1224	gaagtacactgtgagatgtgaa---aggctctgtcgtctatcaatgtgtcctcaagaca	1280
Db	4614	GGGTATACGGGGAGATGGCATTCGGAGACTGTGGAGCTCTTGAGACCCCTCTCTAAGAACAA	4673
OY	1281	tggcggctgcagttccattcttcctcttgcgaactaacactgtgaagaataccaagatatgtac	1340
Db	4674	TGGAGAGTGCAGCCCATATTCACCTGCAAAACACAGGGGATGTGCCAGAGACATGTAC	4733
OY	1341	ctgcaagacgaactaagc---gggtgatgtgaatcgtctcgcggggcagcatctacgcggga	1397
Db	4734	CTTGACACACCCCAACACCCGTGGGGGAGCGCTCACTCTCCCTGCCCCAAGTGGGCTTGA	4793
OY	1398	gcttcccaagaacctctgcagcgtcccaagtaactcttcacagttgtgagagacatgtcgtccg	1457
Db	4794	GCTCCTGAGGATTAAGCATGCTCATTTCTTCAAGCTCCGCGCTCGTGAATAT-----AA	4847
OY	1458	agaagctgtgcagctcgtgcctccatcaacgltgttcgcgcttgtctagtctccatcaatca	1517
Db	4848	GGACTCAAGGCGCATAGGGCTTTTACCAATCTTCTGCGCCGACGAGATCTTAATGAGCA	4907

Qy	1518	tgagccccgattaaagactcggatcaagaaggctcatgtlccaggtcttctgtctata	1577
Db	4908	CCCTGTCAGATATGACGTGGCCGGATTGTGTCGATCGACAGCTGGGTGTCCTACCA	4967
Qy	1578	cgtagtgagctgcgcacagcgtgtctgttgacaacctaaagtgaacaaagtgcacgac	1637
Db	4968	CTGTGGTGGCTGTGCGCGCTGGCGAGCAGAGACTGCTGGAGCAGGGGTATGCGCACGGC	5027
Qy	1638	cctccaaagaaagccagttlccaletctgtlctccaagacatgtgtcataaacaatga	1697
Db	5028	CCTTCAGAGGACACCACTGTGGCTTTCAGCAGAGAGGAGGAGACATATACCTCAATGACTT	5087
Qy	1698	ggcgaagtgctctgtlccaagtacatactcaagcaaatgtgcgtlcaatccagttaaagca	1757
Db	5088	CGCGGGGTGGTGTAGCAGCAGCAATGAGGCGCTGAAAGCGCATCTTCGACCTTATGACCG	5147
Qy	1758	gttgctgtlccccaanaactgtctltaaccccccaagaatgtcctgttgcaaggtlccga	1817
Db	5148	TGTCTGTGTGCCCCCGAAGCGCTGCACGTGGAGGCTGTATGATGCTCCATCCGAGAGAG	5207
Qy	1818	aaacttactaacagtgtgcagcaacaacacagatatccaatcaagcaagtgtatcacga	1877
Db	5208	AAATGTACCCGCGCGCGCCACGAGGCTTCGGTTAACAAGATCTTCAGCGGCTCTGGAAGT	5267
Qy	1878	ctcaaggtctgtctagttcatctaacgtaccatccaacccccagttcaatgtctctggcc	1937
Db	5268	GAGCGGGCTCTGTGCCCCCTGCTTGAAAGAGCATTCCTATAGCCCTTCAAAATGCTTGACC	5327
Qy	1938	taagacaacaagcccttgaagcgtctgtccccagacagcaagacgtcttcgtlcaataaga	1997
Db	5328	CACAGACGCCCTTTTGAGACTCTGCTCCGGATGCGCAGGCTGTGCTACATGAGGA	5387
Qy	1998	caacaagacaagcttgaagtcacttacttgaagttcaagltgacgtgacgtcccaagctlt	2057
Db	5388	CCACCGGTACAAAGTATACACCACTTCTGGGGGGACATATGTCGATGTCGAGGCTT	5447
Qy	2058	agcttgaagactcccceaggtctgtcttccgtgaagaagccctgaaggtctaaggtcgaagt	2117
Db	5448	GGCATCTGACCTGTGCCCAACTGTGGGCCCACTTGAACCAATGAGGAGCCCACTCTCTT	5507
Qy	2118	gaagtgtagaacttgcagtgacatcggttagctcttcttcaacagacaatgtgcagatt	2177
Db	5508	---CTCCGTGACGCCGAGACGGGGCCGGTGAAGCTCATAGTGGGTAGAGTATGCTCGAT	5564
Qy	2178	catacaccgggagactctgttltgaagltggagltggtcctatgtgacatgtgcactactat	2237
Db	5565	TGTGAGGGGCACTTGTGCCCTTTTGAAGGTGGCCGTATGCAATGACACAGCTGTGGA	5624
Qy	2238	gaatctaaccttagtgtgcgaatgtgaacttctaactctgatatccggggagtg	2297
Db	5625	GCCACCTGTGGCTGTGCTCGCTGATGACCACTTGAAGCCCGGCCCTGACACTGAACAC	5684
Qy	2298	cggaaagtgtactt-----tlcaatcccaatggccactgaagagaagcc	2342
Db	5685	CTGCAGCAATCTGTGGCTGAGACGCCACCTGTCTTAGGGGTTACAGAGCAGGACAGCC	5744
Qy	2343	aaaaggc-----gtg	2352
Db	5745	TGAGGCGCTGTGGGGCTTTACCCGGAAGTTCTGAGAGTCCCTCGGTCGACTCTTTGGG	5804
Qy	2353	aagaagaagtgtatatacaaccggttaactcttcaagagaac-----gtgaag	2402
Db	5805	ATTACGAGCGTCTGGGTCCACCCACCACTTTTGGGTAGGGCCCAAGGACCTGTGGAGGG	5864
Qy	2403	ctgcagaacctgtgcaccgltgtlgtalccaacccccagtgctgtcaltacttcat	2462
Db	5865	CTGCCACCGCAATGTGTACACACCACTGGAAGCCACACTCTCTCCCTGTGCACTATGG	5924
Qy	2463	gccaagactgaagctgcgcctgtgaagacacgaataacggttaacaaccgggacatgt	2522
Db	5925	CAGTAGTGCCAACTTGGCTTGCGGAGCCCAAGACCCCTTTATGAGACCGTGGCGGTGG	5984
Qy	2523	ccggaatctgaacacacatgtgaagcgttccattgcacacacgggttcaacggagacg	2582

D	5985	CATGGACGGCATGAGTGGGCAAGTGTCTGTGCCCTTACGCTTTTTCCTGGGACAGC	6044
O	2583	ctcgcagctctgctcgtgcatltygagatlttggcctctgactctcagcccgagctgtccga	2642
D	6045	CTGTGACACTGTGTGCTCTCTGTGTCCCTTTGGGCCCACTTGTCAAGCTCGCCGTGACTGT	6104
O	2643	gcatltygacagltgtatltgatygggagatcacaagctccgcgggagtgctctgtltyaaacaggtg	2702
D	6105	GCAATGCGCGCTGTGATGATAGAGGCGCTTGGGGGCTCTGTGCTCTCTGTGATGAGCGCTG	6164
O	2703	gaacggcgcgtctgtgtgtgacactctccacagctgtatctcgagtgtyacaactgtctgtctc	2762
D	6165	GACTGGGCGCCACGGCTGTGATGAGTGCACACTGGACAGCTGTGTGTACCCACCCCTCTGTC	6224
O	2763	cgtltyacgcacacctgtltyacgagaaacagcgtgtgtgtltaactltyaactltyaagatgtga	2822
D	6225	ACCCGAGAGCTGTGTGCCGTGCAGGCAACACTGTGTAGTGTACACTGTGGCTATTGAAGGGGA	6284
O	2823	cgagatcacatltyacagtgctgtgtgactctltycaaacaaacaaacgggggctgtgcgaagt	2882
D	6285	TGGCGCGGTGTGATGATGAGTGCACACTGTGTCCAGGACGGGATGGTGGCTGATGAGCA	6344
O	2883	cgctaaagtgtctccagaagaagcgcccaagtctctctgcagctltycaagaagaaggtataaagg	2942
D	6345	CGCCAACTGTGAGCAGGATAGGAACAATGTGTCACTTGTACCTCTCTGCGCCGACTACGAGCG	6404
O	2943	ggaatgtctacagctltycaatagatagacccctctgtgacaacgggtgtcaacgggggagltgca	3002
D	6405	TGATGTCTGAGAGCTCGCCGGGCGCCGCAACCCCTGTGCACAGATGGCGCACCGCGGGGCTTGAG	6464
O	3003	tgagcacgcacactltycagaga tga cggggccagcgaagcaagca taaatg tgaatgtaaagaatca	3062
D	6465	CGACGACGCGCAACTGCTTGAGCACACCGGCGCTGAACACACGCGGCTGTGATGCTCCACGAGG	6524
O	3063	ctatgtctcgggagacggagltgagctgtgtgaccttgagcagctgtccg---tcgacgtlgtct	3119
D	6525	CTACGTAGGGGATGAGTGTGCAGTGTGTGAGGAGTGTGAAACCACTGTGTGACACCGCTGCTT	6584
O	3120	acagagaacaagcagagtgagccacacagatctcaactltygaacccctctactltycaagaac	3179
D	6585	GGCGACGACCAACCGCCCTGCGCACTCAATGTCATGTGCACCTGACACGACTTCCAGAGCA	6644
O	3180	gaccgttagagatltccactltaactltaactltaactltygcacgtltygcacagactltyaca	3239
D	6645	ACGGGCTGGGCTTTTCCACTCCAGGCGCAACAGCGGCGCTTATGTGTGAATCTTTCGGA	6704
O	3240	agccaaagaagcctgtgtgccaagaagctgtgacacatagccacatcaaacacagctctcta	3299
D	6705	GGCTGACGGCGGATGCGCAACACAGGAGACCGCTTCTTATTCCTCCACACTCTCTGC	6764
O	3300	tgcccagaagaagcaagatlaactltaactltygtctcggcgcgcgtgtgtgtgagagtgagggtgtgc	3359
D	6765	TGCCCACTAGCTGGGCTTCCACTGTGTGCCATGTGGCTGTGGCTGCATATGCTCTCCACTGC	6824
O	3360	ctaaccgactacgtatgtgctctcagaagtggtgtgtaaaaagtlgtltyggatctgtagacta	3419
D	6825	CCACCGTGTGTTTCCCTGTGTGGCGGACGTGTGGCAATGATGGGTGGGTAGTGAAGCT	6884
O	3420	cgagatcacaggccaacaagaagtgaatgttggaatgtctctgtctaccgatgaagaatgt	3479
D	6885	GGGTGCCCGGAAAGAACTCTCAGAAAGCGTGGGATGCTCACTCTTCGCTGTCAAGATGT	6944
O	3480	gaactltaactltycaaggcgaaggtatlttgggagatgg---cttctgtltyagtggaact	3536
D	6945	GGCTCGGCATATGCCAAATAGCTTTGTGTGGTGCAGGGATACACACTGTCAATGGAAGCACTT	7004
O	3537	gtctcagatctctaatgtctctctccctctgctltycaaaaactcttcagacagagtgctgtctc	3596
D	7005	GCTGGATGTGTGTGCTCCACTGTCCAACTTCTCCACTTCTATGGGATGTCTATTGGCTA	7064
O	3597	ttccaaagatctcagcccgagagacagcatltytgaacaacactltaactltyacatctgtltaacgcg	3656

Db	7065	ttccaatgccaacccagcgggcttctgacttctcttgacttcttgatgtatgagctcacgta	7124
Qy	3657	tggaaccctgtttgtgcacagaacagtgaggctacccgggaataagagcgtgtcgtgcgcg	3716
Db	7125	TAAGACACTCTTGTCGTCCTGTCATATGAAGAGCGCTTGTGTGACAAACATGACGCTGAGTGGCC	7184
Qy	3717	ggacattgagacacacactcactaactatgtcaacgctctcctttaaactgac-----ctt	3768
Db	7185	AGACTTGGAGGCTGATGAGCTTCACAGGCCACCTCTTAAGTGTCCAAAGCCACACGAGGGA	7244
Qy	3769	gtcaatgtgaactcttctgagagactatgcttgggaagcaacgtgcttaacttaactcaagcag	3828
Db	7245	gtttgcttccggcccaactgagcctttagcctcatcattgacgtgacgagcgccttgacaaacag	7304
Qy	3829	gaccagctccacacacagagacca-ggtttgtgtatgaaagatccatctgcagctggagacatg	3887
Db	7305	tttcctggggccctctgctggcccccagagacagtgtgtgtttagccgtatcattgtgtggacat	7364
Qy	3888	categccgacaaatggaatccctccatataattcttgaacctttagagagctctcc	3941
Db	7365	catggcctttcaatgatgcattcatcattgctgtgagccagccctctcgacacccc	7418
RESULT 12			
AX149457			
LOCUS	AX149457	1624 bp	DNA
DEFINITION	Sequence 13 from Patent WO0136638.		linear
ACCESSION	AX149457		
VERSION	AX149457.1	GI:14347934	
KEYWORDS	human.		
SOURCE	Homo sapiens		
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
REFERENCE	1 (bases 1 to 1624)		
AUTHORS	Shimkels, R.A., Lichenstein, H., Vernet, C., and Fernandez, E.		
TITLE	Polypeptides, and nucleic acids encoding same		
JOURNAL	Patent: WO 0136638-A 13 25-MAY-2001;		
FEATURES	Curagen Corporation (US)		
Source	1. 1624 location/Qualifiers		
BASE COUNT	417 a	420 c	386 g 401 t
ORIGIN	/organism="Homo sapiens" /db_xref="taxon:9606"		
Query Match	13.5%;	Score 637.6;	DB 6;
Best Local Similarity	76.0%;	Pred. No. 7.2e-171;	
Matches 830;	Conservative 0;	Mismatches 224;	Indels 38; Gaps 2;
Qy	1508	cttcaatcaatgagcccggaalttaagacttggaatcagcaagagcgtctatgtcccaaggttc	1567
Db	571	CTCTGATCTTCTGTGTAAAGGTTAAAGCTGGGACAAATACGGTTTATGCCCCAGGTTTC	630
Qy	1568	ttcgtatcacgctggtggtgtcgcagcagctgtgtcttgcacaacctlaaagtgacaccaa	1627
Db	631	tttcggtaccatgtgtgtcccttcctccacacgctgtcttcttgaaaacctgaaattgattctaa	690
Qy	1628	gtgccagagcccttccaagagagagccagttlccatcctgtcctcctcagacacgtgtgtca	1687
Db	691	ATGCTACTTCCCTCCCAAGAGAGCACAATAGTATCTCCGCTCTCAGACACGGGTATTA	750
Qy	1688	taaacaaatgagagcgaaggttcgtctcattgatcatatataagacaaccaatgtggtcalccagc	1747
Db	751	TAAACAAATPAGGCTTAAATGATCATATCATGTGATATCATAGTACTAATVGGATTGTTCTA	810
Qy	1748	ttatagacaagtgctgtcttcccaaaacttgcttatcacccccaaagatgcttggaca	1807
Db	811	TCATAGACAAATGCTATATCTCCCAAAATTTGCTTATATCATTCCCAAAACAACTGTGAA	870
Qy	1808	gggttctgcaaaacttactacatgctgacgaacacacgatalacccaattcagcaagt	1867

```

Db      871 GAATTCGCAAAATCTTACGACTTTGGCAACAAACAATGGCTACATCAATTAATGCAACT 930
      1868 tgatagaagactaaggcttgcgtcgaactcaactgaactcaccacacccagtcactg 1927
      931 TAATACAGGACTAGGTTTCTGAGTGTCAACCGATCCCATCCACACCCAGTCACTC 990
      1928 tctcttgcccttagcagacaagcccttggaagcccttgcccccaagcagcaggaactctgt 1987
      991 TCTTTGGCCACGACGACCAAGCCCTCCATGCCCTACTGCTGTAACAACAGACACTTCTGT 1050
      1988 tcaatcaagaacacaagaagaagaactggaagcttcaactggaagttccaaagttaccgaagact 2047
      1051 TCAACCAACCAACCAACCAACCAAGCTGAAGATTTGATTTGATTTGATTTGATTTGATTTG 1110
      2048 ccaagagcttagcttagcactcccccaggtcgtgcttccctggaagaccctggaagctcag 2107
      1111 CCAAGGTTTAACTGTGTGATCTTCCACATCCACTGCTGGAACACCCCTGCAAGTTTCAG 1170
      2108 agctgaagtgtgaaggtgtggaagacttgacatcagtgacatcgttcttctaaacgaacaaa 2167
      1171 AGCTGAGTGTGAATGTGAGACTGGACGGACATCGGATCGGACTCTTCTGTAATGGCCAAA 1230
      2168 tctgcagatcacaacacccgggaactctgttgcagctggtgtgtgctcctatgacatgact 2227
      1231 CCTGCAAAATGTGTCAACGCGGAGCTCTTGTGACTGTGGGTGTGGCTTACGGCAATGACT 1290
      2228 gctcactcactgaactcactcactgaagtgccgaatgtagacacttactactcactgatalc 2287
      1291 GTCTGCTGATTTGATTTCCACCTTGGGGGGCGCTGTGACACTTTTACTTACTTCTGATGCT 1350
      2288 cggggagagtcggaagtgatgtacatttccacaaatgcccacatgaaagaagaacaaagg 2347
      1351 CGGGGAGATGTGGAGGCTGTGTCAATACTCCACACTGCCAAGGTGGAGTAACCAAAAGG 1410
      2348 gctgtgaagaagaagtgtatctcaaacccgttacccttccagaaggaagtggaaggtgccc 2407
      1411 GTTGAACCAACCAAGTGTCTTACCAACC--TGCCCTTTAAAGAAAGAACTTGAAGAGCTGCC 1467
      2408 agaaccctgtgacccgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2467
      1468 GGGAGCGGTGCAAGCTGTGTGTATACGATCCCA----- 1500
      2468 actgtcaagcctgtcccttggaagacagatacaccgtgttaacaacccgggacatgtgcgcg 2527
      1501 -----GCTGCCCTGAGAGACCAAGATGCCCTGTAAATAACCGGGGTGTCTGCTTGG 1552
      2528 atcgttacaacacatgtggaacagtgctatgtccacacccggttcaacgggaagacgtcgcg 2587
      1553 ATCACTACTCGGCGCACCGAGAGTGTAAATGCAACACCGGCTTCAATGGAGCGGCTGTG 1612
      2588 agctctgtctgc 2599
      1613 AGATGTCTGTGC 1624

```

RESULT 13
AX149463
LOCUS AX149463 1577 bp DNA linear PAT 08-JUN-2001
DEFINITION Sequence 19 from Patent WO0136638.
ACCESSION AX149463
VERSION AX149463.1 GI:14347937
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 (bases 1 to 1577)
 Shimkets,R.A., Lichenstein,H., Vernet,C. and Fernandes,E.
TITLE Polypeptides and nucleic acids encoding same
JOURNAL Patent: WO 0136638-A 19 25-MAY-2001;
 Curagen Corporation (US)
FEATURES
 Location/Qualifiers
 source 1..1577

```

BASE COUNT      424 a      410 c      342 g      401 t
ORIGIN
Query Match      11.0%; Score 517.4; DB 6; Length 1577;
Best Local Similarity 78.8%; Pred. No. 2e-136;
Matches 617; Conservative 0; Mismatches 166; Indels 0; Gaps 0;

      1508 ccttcaatcatalgaagcccggaaltaaagacttggatcagcagagggcctcatgtlcccaagttc 1567
      571 CCTGTGATCTTCTGTGTACAGTTTAAACACGTGGGACCAATACGGTTTATGGCCAGGTTTC 630
      1568 ttgcatacaagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1627
      631 TTCGGTACCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 690
      1628 gtgcacagaccccttcaagggagccagtttccatctctgtctcctcaagacatgtgttca 1687
      691 ATGCTACTTCCCTCCAGGAGAGCAATATGATCTCTGCTCTCAAGACAGGTTGATA 750
      1688 taaacaatgaagcggaaggtcctgtccagtgacatcatcagacccaatgtgcatacagc 1747
      751 TAAACATTAAGGCTAAGATCATATCCACTGATATCATGATCAATGAATGGATGTTCATA 810
      1748 ttatagaagaatgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1807
      811 TCATAGAACAAATGTCTACTATCCCAAAATTTGTATCATCTCCCAAAAGCAACACTGTGAA 870
      1808 ggtgttctgcaaatcttactcagtggtgcaacaacccaagagatatataccaatctcaagaagt 1867
      871 GAATTCGCAAAATCTTACGACTTTGGCAACAAACAATGGCTACATCAATTAATGCAACT 930
      1868 tgataagaagactcaggtctgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 1927
      931 TAATACAGGACTCAGGTTTGTGAGTGTCAACCGATCCCATCCACACCCAGTCACTC 990
      1928 tctcttgcccttagcagacaagcccttggaagcccttgcccccaagcagcaggaactctgt 1987
      991 TCTTTGGCCACGACGACCAAGCCCTCCATGCCCTACTGCTGTAACAACAGACACTTCTGT 1050
      1988 tcaatcaagaacacaagaagaagaactggaagcttacccttgaagttccacagtgatccagact 2047
      1051 TCAACCAACCAACCAACCAACCAAGCTGAAGAGTATTTAAGTTTATGTGATACGAGATG 1110
      2048 ccaagcctttagcctcagaacctcccaagctgtccttcccttggaagaacccctgcaaggtcag 2107
      1111 CCAAGGTTTAACTGTGTGATCTTCCACATCCACTGCTGGAAGACCCCTGCAAGGTTTCAG 1170
      2108 agctgaagtgtgaaggtgtggaacttgacatcagtcgtgtgaagctcttctttaaagaaacaa 2167
      1171 AGCTGAGTGTGAATGTGAGAGCTGTGACGAGGACATGGTGACCTCTTGTGAAATGGCCAAA 1230
      2168 tgtgcagatcacaacacccgggaactcgtgttgaagtggtgtgtgtgtgtgtgtgtgtgtgt 2227
      1231 CCTGCAGAATTTGTCAAGCGGAGCTTGTGTTGACTTGAGCTGGGTTGGCTTACGGCATTTGACT 1290
      2228 gctactcaatgaactcactcactgaagtggtgcgaatgtgaacatttactactcagataltc 2287
      1291 GTCTGCTGATTTATCCACACCCGCGGCGCTGTGACACCTTTACTACTTTCGATGCT 1350
      2288 cgg 2290
      1351 CGG 1353

```

RESULT 14
AC025501
LOCUS AC025501 218073 bp DNA linear HTG 25-JUN-2001
DEFINITION Mus musculus chromosome 10 clone RP23-129023 map 10, WORKING DRAFT
SEQUENCE 18 unordered pieces.

VERSION AC025501.4 GI:14547868
 KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 218073)
 AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 TITLE Mus musculus chromosome 10, clone RP23-129023
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 218073)
 AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,
 Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F.,
 Boguski, J., Boukhalter, B., Brown, A., Burkett, G.,
 Campiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,
 Collymore, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, D.S.,
 Dodge, S., Domino, M., Doyle, M., Ferrel, P., Fitzhugh, W., Gage, D.,
 Galagan, J., Gardina, S., Ginde, S., Goyette, M., Graham, L.,
 Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L.,
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,
 Klein, J., Larocque, K., Lamazares, R., Landers, T., Lebeck, J.,
 Levine, R., Lieu, C., Liu, G., Locke, K., MacDonald, P., Marquis, N.,
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheters, R.,
 Meldrum, J., Menus, L., Mihova, T., Miranda, C., Mlenka, V., Morrow, J.,
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,
 O'Neill, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rotman, D.,
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,
 Teste, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,
 Vassiliev, H., Viel, R., Vo, A., Wilson, R., Wu, X., Wyman, D., Ye, M.J.,
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.
 COMMENT All repeats were identified using RepeatMasker:
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 TITLE JOURNAL
 COMMENT
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: W1BR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 Project Information
 Center project name: 129_O-23
 Center clone name: 129_O-23
 Summary Statistics
 Sequencing vector: M13; M77815; 44% of reads
 Sequencing vector: Plasmid; n/a; 56% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 213014 bases at least Q40
 Consensus quality: 214764 bases at least Q30
 Consensus quality: 215528 bases at least Q20
 Insert size: 210000; agarose-fp
 Insert size: 216373; sum-of-coverage
 Quality coverage: 10.2 in Q20 bases; agarose-fp
 Quality coverage: 9.9 in Q20.
 NOTE: This is a 'working draft' sequence. It currently
 consists of 18 contigs. The true order of the pieces
 is not known and their order in this sequence record is
 arbitrary. Gaps between the contigs are represented as
 runs of N, but the exact sizes of the gaps are unknown.
 This record will be updated with the finished sequence
 as soon as it is available and the accession number will
 be preserved.
 1 23321: contig of 23321 bp in length
 * 23322 23421: gap of 100 bp
 * 23422 24529: contig of 1108 bp in length
 * 24530 24629: gap of 100 bp
 * 24630 26959: contig of 2330 bp in length
 * 26960 27059: gap of 100 bp

FEATURES
 source
 * 27060 28240: contig of 1181 bp in length
 * 28241 28340: gap of 100 bp
 * 28341 30420: contig of 2080 bp in length
 * 30421 30520: gap of 100 bp
 * 30521 32131: contig of 1611 bp in length
 * 32132 32231: gap of 100 bp
 * 32232 34170: contig of 1939 bp in length
 * 34171 34270: gap of 100 bp
 * 34271 36716: contig of 2446 bp in length
 * 36717 36816: gap of 100 bp
 * 36817 40349: contig of 3533 bp in length
 * 40350 40449: gap of 100 bp
 * 40450 43365: contig of 2916 bp in length
 * 43366 43465: gap of 100 bp
 * 43466 46874: contig of 3409 bp in length
 * 46875 46974: gap of 100 bp
 * 46975 51691: contig of 4717 bp in length
 * 51692 51791: gap of 100 bp
 * 51792 118666: contig of 66875 bp in length
 * 118667 118766: gap of 100 bp
 * 118767 133750: contig of 14984 bp in length
 * 133751 133850: gap of 100 bp
 * 133851 151236: contig of 17366 bp in length
 * 151237 151336: gap of 100 bp
 * 151337 174475: contig of 23139 bp in length
 * 174476 174575: gap of 100 bp
 * 174576 208697: contig of 34122 bp in length
 * 208698 208797: gap of 100 bp
 * 208798 218073: contig of 9276 bp in length.
 Location/Qualifiers
 1. 218073
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="10"
 /map="10"
 /clone="RP23-129023"
 /clone_1bp="RPC1-23 Female Mouse BAC"
 1. 23321
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left"
 23422. 24529
 /note="assembly_fragment"
 24630. 26959
 /note="assembly_fragment"
 27060. 28240
 /note="assembly_fragment"
 28341. 30420
 /note="assembly_fragment"
 30521. 32131
 /note="assembly_fragment"
 32232. 34170
 /note="assembly_fragment"
 34271. 36716
 /note="assembly_fragment"
 36817. 40349
 /note="assembly_fragment"
 40450. 43365
 /note="assembly_fragment"
 43466. 46874
 /note="assembly_fragment"
 46975. 51691
 /note="assembly_fragment"
 51792. 118666
 /note="assembly_fragment"
 118767. 133750
 /note="assembly_fragment"
 133851. 151236
 /note="assembly_fragment"
 151337. 174475
 /note="assembly_fragment"
 174576. 208697
 /note="assembly_fragment"


```
*      87869      91230: contig of 3362 bp in length
*      91231      91330: gap of unknown length
*      91331      95333: contig of 4003 bp in length
*      95334      95433: gap of unknown length
*      95434      10081: contig of 4648 bp in length
*      10082      10081: gap of unknown length
*      10082      102921: contig of 2740 bp in length
*      10292      103021: gap of unknown length
*      10302      107381: contig of 4260 bp in length
*      10738      107382: gap of unknown length
*      107382      110228: contig of 2847 bp in length
*      11022      110328: gap of unknown length
*      11032      114370: contig of 4042 bp in length
*      11437      114471: gap of unknown length
*      11447      117842: contig of 3372 bp in length
*      11784      117943: gap of unknown length
*      11794      122460: contig of 4518 bp in length
*      12246      122560: gap of unknown length
*      12256      126561: contig of 4080 bp in length
*      12656      126740: gap of unknown length
*      12674      130890: contig of 4150 bp in length
*      13089      130990: gap of unknown length
*      13099      134751: contig of 3761 bp in length
*      13475      134851: gap of unknown length
*      13485      138027: contig of 3176 bp in length
*      13802      138127: gap of unknown length
*      13812      141177: contig of 3050 bp in length
*      14117      141278: gap of unknown length
*      14127      143823: contig of 2546 bp in length
*      14382      143923: gap of unknown length
*      14392      146573: contig of 2650 bp in length
*      14657      146673: gap of unknown length
*      14667      149832: contig of 3159 bp in length
*      14983      149933: gap of unknown length
*      14993      152476: contig of 2538 bp in length
*      15247      152570: gap of unknown length
*      15257      155089: contig of 2519 bp in length
*      15509      155189: gap of unknown length
*      15519      157216: contig of 2027 bp in length.
```

FEATURES

```
Source      Location/Qualifiers
            1..157216
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="3"
            /clone="RP11-54J3"
```

```
BASE COUNT      44517 a 32787 c 33678 g 42931 t 3303 others
ORIGIN
```

```
Query Match      3.1%: Score 146.8: DB 2: Length 157216:
Best Local Similarity 83.8%: Pred. No. 5.1e-30:
Matches 166: Conservative 0: Mismatches 32: Indels 0: Gaps 0:

QY 2837 cagtcgtgactctgcacaacagaacaaggggctgtgcgaagtlcctaagtctccc 2896
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114169 CAGTTGTGATTTTGCACAAACAGGACAGGGGGCTGTGCAAGAGTGCCAGATCTCCC 114110

QY 2897 agaaagcaccacgaagtccttcgagctgcgaagaagctcacaaaggggatatgctacagct 2956
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114109 AGAAGGGCAGCAGAGCTCTCCAGAGCTCCAGAGGGATACAAAGGGGACGCGCACAGCT 114050

QY 2957 gcatagagatagaccctgttcacagcgtgtcacaagggggtatgcataagacggcacct 3016
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 114049 GCACAGAGATGAGACCCCTGTCTCAGAGCGCTTAAAGAGGGTGTACGACGACGCGCACCT 113990

QY 3017 gcagatgacgggcccag 3034
      | | ||| ||| ||| |
Db 113989 GTAAGATGACAGGCCCGG 113972
```

Search completed: June 17, 2002, 16:45:43
Job time: 12758 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 14:00:15 : Search time 465.74 Seconds
(without alignments)
17348.305 Million cell updates/sec

Title: US-09-842-930A-1
Perfect score: 4706
Sequence: 1 tctttaccgaactactacac.....aagcaaaaaaaaaaaaaa 4706

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: _N_Geneseq_032802:*
2: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1980.DAT:*
3: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1981.DAT:*
4: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1982.DAT:*
5: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1983.DAT:*
6: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1984.DAT:*
7: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1985.DAT:*
8: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1986.DAT:*
9: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1987.DAT:*
10: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1988.DAT:*
11: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1989.DAT:*
12: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1990.DAT:*
13: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1991.DAT:*
14: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1992.DAT:*
15: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1993.DAT:*
16: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1994.DAT:*
17: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1995.DAT:*
18: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1996.DAT:*
19: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1997.DAT:*
20: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/hold-geneseq/geneqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4706	100.0	4706	24	ABA04648
2	2835.2	60.2	4576	24	ABA04662
3	1599.4	34.0	3625	23	AAF87120
4	1244.8	26.5	2483	23	AAF87119
5	1243.2	26.4	2512	23	AAF87117
6	1196	25.4	2011	23	AAF87113
7	1016.4	21.6	1804	23	AAF87114
8	883.4	18.8	1259	21	AAA57365
9	695.6	14.8	6761	21	AAA57362

10	654.4	13.9	1377	21	AAC76373	Human ORFX ORF1928
11	637.6	13.5	1624	23	AAF87118	NOV7 coding sequen
12	517.4	11.0	1577	23	AAF87121	NOV7 coding sequen
13	220.8	4.7	1522	21	AAAS7363	DNA encoding a hum
14	177.4	3.8	718	22	AA500854	Human cDNA clone H
15	109.4	2.3	722	24	AA562772	cDNA sequence #559
16	70.8	1.5	1414	13	AAQ27190	Tumour necrosis fa
17	70.8	1.5	1414	13	AAV71778	Tumour necrosis fa
18	70.8	1.5	1414	22	AAQ60319	Human tumour necro
19	70.2	1.5	1182	23	AA584364	DNA encoding novel
20	63	1.3	1411	22	AAH23114	Osteoarthritis tis
21	57.4	1.2	1354	22	AAQ08974	Alternative versio
22	56.6	1.2	985	21	AAAS7364	DNA encoding a hum
23	56.6	1.2	1365	22	AA158121	Human polynucleoti
24	56.6	1.2	1984	20	AAAS2252	Protein PRO271 CDN
25	56.6	1.2	1985	22	AAAS5938	Human DNA encoding
26	56.6	1.2	1985	22	AAAF2410	Human PRO271 cDNA.
27	55.4	1.2	781	22	ABA08712	Human stabilin-I h
28	55.4	1.2	2808	21	AAAS9054	Human secreted pro
29	52.6	1.1	1851	22	ABA83182	HGST-4 ovarian tum
30	51.2	1.1	11344	23	AA592464	DNA encoding novel
31	49.8	1.1	4480	19	AAV43590	Human epithelial g
32	49.8	1.1	4877	22	AAH28219	Nucleotide sequenc
33	49.8	1.1	5007	22	AAK51657	Human polynucleoti
34	49.8	1.1	5032	22	ABA09190	Human precursor pr
35	49.8	1.1	5032	22	AAK52641	Human polynucleoti
36	49.6	1.1	6567	11	AAQ03324	Etmeria tenella ge
37	45.4	1.0	1422	23	AA592463	DNA encoding novel
38	44.2	0.9	2895	22	AA167200	Nucleotide sequenc
39	44.2	0.9	2897	22	AAH78221	Human bone marrow
40	42.8	0.9	436	22	AAK30592	Human secreted pro
41	42.2	0.9	1308	21	AAAS1263	Human cDNA SEQ ID
42	42.2	0.9	1640	22	ABA65683	Human immunoglobul
43	42.2	0.9	1641	22	AAAS2787	Human cDNA SEQ ID
44	42.2	0.9	1641	22	ABA60754	Human immunoglobul
45	42.2	0.9	1641	22	AA528848	Human immunoglobul

ALIGNMENTS

RESULT 1	ABA04648	standard; cDNA: 4706 BP.
ID	ABA04648	
AC	ABA04648:	
XX		
DT	22-FEB-2002	(first entry)
XX		
DE	Rat Hyaluronon Acid Receptor for Endocytosis, HARE, coding sequence.	
XX		
KW	HARE, Hyaluronon Acid Receptor for Endocytosis; hyaluronan; chondroitin;	
KW	chondroitin sulphate; extracellular matrix; cartilage; skin;	
KW	vitreous humour; endocytic receptor; glycosaminoglycan; rat; ss.	
XX		
OS	Rattus norvegicus.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..4296
FT		/*tag= a
FT		/partial
FT		/product= "Rat HARE"
FT		/note= "No start codon given"
XX		
PN	WO200181544-A2.	
XX		
PD	01-NOV-2001.	
XX		
PE	25-APR-2001; 2001WO-US13403.	
XX		
PR	25-APR-2000; 2000US-199538P.	
XX		
PR	02-NOV-2000; 2000US-245320P.	
XX		

Db 1681 gfttcataaacaatgagcggaagttccgtgccagtacatcatcagccaatgycgtc 1740

QY 1741 atccacgtatagacaagtgtgtctccccaanaacttgcttatcaacccccaaagtgc 1800

Db 1741 atccacgtatagacaagtgtgtctccccaanaacttgcttatcaacccccaaagtgc 1800

QY 1801 ttggcagaagtctctgcacaatatcttaacagtgagcagcaaacacagatataccaatlc 1860

Db 1801 ttggcagaagtctctgcacaatatcttaacagtgagcagcaaacacagatataccaatlc 1860

QY 1861 agcaaatgtatagacaaggaactcaagcttgctgtcagtcacatcaactgaactccacacca 1920

Db 1861 agcaaatgtatagacaaggaactcaagcttgctgtcagtcacatcaactgaactccacacca 1920

QY 1921 gtcactgtctcttgctgccttaagcgaagcccttggaagccttgcccccaagacagcagac 1980

Db 1921 gtcactgtctcttgctgccttaagcgaagcccttggaagccttgcccccaagacagcagac 1980

QY 1981 ttccgttcaatcaagaacaagaacaagcctgaaagtccttaacctgaagttccaagtgatc 2040

Db 1981 ttccgttcaatcaagaacaagaacaagcctgaaagtccttaacctgaagttccaagtgatc 2040

QY 2041 cgaagactccaaggtcttaagcttcaagactccccaagctctgtcttccgtgaaagccctgc 2100

Db 2041 cgaagactccaaggtcttaagcttcaagactccccaagctctgtcttccgtgaaagccctgc 2100

QY 2101 ggtcagaagcttgagtgtagtggtggaacttgcaatgacatccggtgagcttcttccaaac 2160

Db 2101 ggtcagaagcttgagtgtagtggtggaacttgcaatgacatccggtgagcttcttccaaac 2160

QY 2161 gaacaatatgtgcagatcatatacaccgggagactctgtttgaaagtgtgttgacctatgac 2220

Db 2161 gaacaatatgtgcagatcatatacaccgggagactctgtttgaaagtgtgttgacctatgac 2220

QY 2221 attgactgtcactcatatgaatcccttaagcttgaagcttgagcttggtgtgacctatgac 2280

Db 2221 attgactgtcactcatatgaatcccttaagcttgaagcttgagcttggtgtgacctatgac 2280

QY 2281 gatattccgggggagtgcggaagtgtgcatcttaccctcccaatgtgccaatgaaaggaag 2340

Db 2281 gatattccgggggagtgcggaagtgtgcatcttaccctcccaatgtgccaatgaaaggaag 2340

QY 2341 ccaaaagggcgtgaaagaagaagtgtatcttaacaacccgttacccttcaaggaagaatgaa 2400

Db 2341 ccaaaagggcgtgaaagaagaagtgtatcttaacaacccgttacccttcaaggaagaatgaa 2400

QY 2401 ggcgtgcgaagaacctgtgcacgctgtgtatccaaccccccaagtgctgccaatgttacttc 2460

Db 2401 ggcgtgcgaagaacctgtgcacgctgtgtatccaaccccccaagtgctgccaatgttacttc 2460

QY 2461 atgccaagactgtcagagccctgagccttgaggaacagatacaccgtgtgtaaacaccgggcatg 2520

Db 2461 atgccaagactgtcagagccctgagccttgaggaacagatacaccgtgtgtaaacaccgggcatg 2520

QY 2521 ttgcgcgatactgttaacaacccaatgggacagtgctatactgcacaacggcttcaacgggaca 2580

Db 2521 ttgcgcgatactgttaacaacccaatgggacagtgctatactgcacaacggcttcaacgggaca 2580

QY 2581 gctctgcagagctctgtcgtgcacatgggagatttggtccttgactgtcagccccgcagctgtccc 2640

Db 2581 gctctgcagagctctgtcgtgcacatgggagatttggtccttgactgtcagccccgcagctgtccc 2640

QY 2641 gagcattgacaagtgtgataagagggaatcaacaggtcccgagggaatgtgctctgtgtaaacagg 2700

Db 2641 gagcattgacaagtgtgataagagggaatcaacaggtcccgagggaatgtgctctgtgtaaacagg 2700

QY 2701 ttggaagccgctctcgtgtgtacaactcccaagctgtatattcgacagtgtgcacaactgtctgc 2760

Db 2701 ttggaagccgctctcgtgtgtacaactcccaagctgtatattcgacagtgtgcacaactgtctgc 2760

QY 2761 tccgtgcagcgcacactgtgatacgggaagaacaacagctgtgtgtgttaacttgaactcgaaggt 2820

Db 2761 tccgtgcagcgcacactgtgatacgggaagaacaacagctgtgtgtgttaacttgaactcgaaggt 2820

QY 2821 gacgggatacatatgcacagtcgttgacactctgcacaacagaacaacaggggctgtgcgaag 2880

Db 2821 gacgggatacatatgcacagtcgttgacactctgcacaacagaacaacaggggctgtgcgaag 2880

QY 2881 gtcgcataagtctcccaagaaagcaccacaagtctcttgcaagtgtgcagaagaagcttaacag 2940

Db 2881 gtcgcataagtctcccaagaaagcaccacaagtctcttgcaagtgtgcagaagaagcttaacag 2940

QY 2941 ggggatagtctacagctgtcaatagaatagaacccctgtgcagacagtgltcaaggggatgac 3000

Db 2941 ggggatagtctacagctgtcaatagaatagaacccctgtgcagacagtgltcaaggggatgac 3000

QY 3001 catgtgacaagccacccctgcagagatgacagggccccaggaacagatgaatgtgaatgtaaagt 3060

Db 3001 catgtgacaagccacccctgcagagatgacagggccccaggaacagatgaatgtgaatgtaaagt 3060

QY 3061 caataatgtcgggagagaaatgagactgtgagcctggaacagctgcgcctgcacccgttgatca 3120

Db 3061 caataatgtcgggagagaaatgagactgtgagcctggaacagctgcgcctgcacccgttgatca 3120

QY 3121 caggaacaacgagacagtgaccacccagatgcccagctgtgcagacctacttccaagaacag 3180

Db 3121 caggaacaacgagacagtgaccacccagatgcccagctgtgcagacctacttccaagaacag 3180

QY 3181 accgtgagagatacttccatctacgctcccaactgggcaggtacaaatgacatttgaca 3240

Db 3181 accgtgagagatacttccatctacgctcccaactgggcaggtacaaatgacatttgaca 3240

QY 3241 gccaaagaagcctgtgcacaagaagctgacacatatagccacacttaacaacagactctccat 3300

Db 3241 gccaaagaagcctgtgccaagaagctgacacatatagccacacttaacaacagactctccat 3300

QY 3301 gcccaagaagcccaagtatacactgtgtcgtgcgcggctgagctgagagatgagcggtgtgcc 3360

Db 3301 gcccaagaagcccaagtatacactgtgtcgtgcgcggctgagctgagagatgagcggtgtgcc 3360

QY 3361 taacccgactagatgtgcctctcaagaagtgtgtgcacaagctgtgttgagatctgatactac 3420

Db 3361 taacccgactagatgtgcctctcaagaagtgtgtgcacaagctgtgttgagatctgatactac 3420

QY 3421 ggaatccagggccaacaagaatgtaaatgttggaatgtcttctgttaccggaatgaagaatgtg 3480

Db 3421 ggaatccagggccaacaagaatgtaaatgttggaatgtcttctgttaccggaatgaagaatgtg 3480

QY 3481 aactgtcacctgcagaagcaggtcatatgttgagagatgtgcttctcgttgcaagtggaaactgtgc 3540

Db 3481 aactgtcacctgcagaagcaggtcatatgttgagagatgtgcttctcgttgcaagtggaaactgtgc 3540

QY 3541 caggttccctcatgtcttccctccgtccacaaacttccctgacagaaggtgtgtgcttcttcc 3600

Db 3541 caggttccctcatgtcttccctccgtccacaaacttccctgacagaaggtgtgtgcttcttcc 3600

QY 3601 aagaactcagcccgagggaagaagcatcttctgaacaacacttcaacagactgtccatccgtgc 3660

Db 3601 aagaactcagcccgagggaagaagcatcttctgaacaacacttcaacagactgtccatccgtgc 3660

QY 3661 accctgttctgtgcacaagaacaagtgggtctaacgggaaataaagaagctgtgtgcgggagac 3720

Db 3661 accctgttctgtgcacaagaacaagtgggtctaacgggaaataaagaagctgtgtgtgcgggagac 3720

QY 3721 atttgacacacactcaactaatgttcaagctctcctttaaactgaagactgttcaatgtgttac 3780

Db 3721 atttgacacacactcaactaatgttcaagctctcctttaaactgaagactgttcaatgtgttac 3780

QY 3781 ttcttgagactatgtctgggaagccaactgtcattacacttcaagccaggaacagcttccac 3840

Db 3781 ttcttgagactatgtctgggaagccaactgtcattacacttcaagccaggaacagcttccac 3840

QY 3841 caagagacaaagtctgttggaatgaagaatccatctctgcagtgagacataatgcgcgcaat 3900

Db 3841 caagagacaaagtctgttggaatgaagaatccatctctgcagtgagacataatgcgcgcaat 3900

QY	3901	ggatctctccataattatttcttgaaccttttgagagctccctccacagcgaacaacggctgccc	3960
Db	3901	ggaaacctccatatattatcttgaaccttttgagagcctccctccacagcgaacaacggctgccc	3960
QY	3961	caactctgacctggggagcaaggtatattctgctgcccgtctgcctctgtgcaactctggtgcatgtct	4020
Db	3961	caactctgacctggggagcaaggtatattctgctgcccgtctgcctctgtgcaactctggtgcatgtct	4020
QY	4021	cttggcagcttaactcttaactctcgctcaaaagcagcgaaacctggtttccagcgctttgatt	4080
Db	4021	cttggcagcttaactcttaactctcgctcaaaagcagcgaaacctggtttccagcgctttgatt	4080
QY	4081	cagaagaaggagcatlgtatgtctcttcttgcagaagcagcagccccaaagaatattgcgcaaac	4140
Db	4081	cagaagaaggagcatlgtatgtctcttcttgcagaagcagcagccccaaagaatattgcgcaaac	4140
QY	4141	ctctgtatgtgagactctcagcgcgcgcgaccccccaagttccctctgtgtgacctcttcaagacc	4200
Db	4141	ctctgtatgtgagactctcagcgcgcgcgaccccccaagttccctctgtgtgacctcttcaagacc	4200
QY	4201	cttgagagacacggagatcttgagagacagcagacctcttggggagactgcggtctctgacatgaga	4260
Db	4201	cttgagagacacggagatcttgagagacagcagacctcttggggagactgcggtctctgacatgaga	4260
QY	4261	agccagaagaagcaaccacagttcacggttccacggttgatctccagcccccaagctgtctcatg	4320
Db	4261	agccagaagaagcaaccacagttcacggttccacggttgatctccagcccccaagctgtctcatg	4320
QY	4321	atcagtggtttttaagaatgaaacaactctaaagcagcagcacaactcctcaacctcttggtta	4380
Db	4321	atcagtggtttttaagaatgaaacaactcacaactctaaagcagcagcacaactcctcaacctcttggtta	4380
QY	4381	atccggagattctgcgcagcggctaaagcgcatgttgccttgatataccttggggagacctccac	4440
Db	4381	atccggagattctgcgcagcggctaaagcgcatgttgccttgatataccttggggagacctccac	4440
QY	4441	ctccctctgagccttataccgtggttctctcctcaactccacatagtgtgctggtctgtcttcgccc	4500
Db	4441	ctccctctgagccttataccgtggttctctcctcaactccacatagtgtgctggtctgtcttcgccc	4500
QY	4501	tctctcttgtaaccacaacactgtagactctgtgtgattctctccattgagctgaagcaccacaag	4560
Db	4501	tctctcttgtaaccacaacactgtagactctgtgtgattctctccattgagctgaagcaccacaag	4560
QY	4561	cggggcttcacaccttlatgtctgtattccagttaccagaagatcaacctgccaacatgtgt	4620
Db	4621	gctcaataaagtgttttggaacaataaagaagaagcactgtgtaacctagaaggtgtcaaa	4680
QY	4621	gctcaataaagtgttttggaacaataaagaagaagcactgtgtaacctagaaggtgtgtcaaa	4680
Db	4681	ctatgaaagcagaaaaaataaaaaaaa 4706	
	4681	ctatgaaagcagaaaaaataaaaaaaa 4706	
RESULT 2			
ABA04662			
ID	ABA04662	standard; cDNA: 4576 BP.	
XX	AC	ABA04662;	
XX	DT	22-FEB-2002 (first entry)	
DE	XX	Human Hyaluronic Acid Receptor for Endocytosis, HARE, coding sequence.	
XX	KX	HARE: Hyaluronic Acid Receptor for Endocytosis; hyaluronan; chondroitin;	
KW	KM	chondroitin sulphate; extracellular matrix; cartilage; skin;	
XX	KX	vitreous humour; endocytic receptor; glycosaminoglycan; human; ss.	
XX	OS	Homo sapiens.	
XX	XX		

Key	Location/Qualifiers
FT CDS	1..4185
FT	/**tag= a
FT	/partial
FT	/product= "Human HARE"
FT	/note= "No start codon given"
PN	MOZ00181544-A2.
XX	
PD	01-NOV-2001.
XX	
PF	25-APR-2001; 2001MO-USJ3403.
XX	
PR	25-APR-2000; 2000US-199538P.
XX	
PR	02-NOV-2000; 2000US-245320P.
XX	
PA	(WEIG/) WEIGEL P A.
XX	
PA	(ZHOU/) ZHOU B.
XX	
PA	(WEIG/) WEIGEL J A.
PI	Weigel PA, Zhou B, Weigel JA:
DR	WPI: 2002-049271/06.
XX	
DR	P-PSDB: AAM47684.
XX	
PT	New mammalian hyaluronic acid receptor for endocytosis, useful e.g. for
PT	identifying agents that inhibit binding to hyaluronic acid, and related
PT	nucleic acid -
XX	
PS	Claim 49; Fig 33; 263pp: English.
XX	
CC	The present invention relates to sequences for rat and human HARE
CC	(Hyaluronic Acid Receptor for Endocytosis, ABA04648, ABA04662, AAM47675
CC	and AAM47684). HARE can bind specifically to at least one of hyaluronidic
CC	acid (HA, also known as hyaluronan), chondroitin (C) or chondroitin
CC	sulphate (CS). HA is an extracellular matrix component of all tissues,
CC	in particular cartilage, skin and vitreous humour. HARE is the endocytic
CC	receptor responsible for removing HA and other glycosaminoglycans from
CC	the circulation. The present sequence is the coding sequence for human
CC	HARE.
CC	
CC	
CC	
SQ	Sequence 4576 BP; 1127 A; 1193 C; 1203 G; 1053 T; 0 other;
Query Match	60.2%; Score 2835.2; DB 24; Length 4576;
Best Local Similarity	79.2%; Pred. No. 0;
Matches 3483; Conservative	0; Mismatches 898; Indels 19; Gaps
Oy	67 attcatatacaactgagcaagtgcgaatcgagctgtgcagatgcttatactgttttcgcca 126
Dd	60
Oy	1 attcataataatttgccgatggaaattggagtggttcgccagatgcttaacagatgtttgtcca 60
Dd	60
Oy	127 aacaatgaagaccatgcgaaaactataatcaggagagaagaagccacatctctaagaagaat 186
Dd	120
Oy	61 aacaaacaatgcccattcgagaatatcacatcaggagagaagaagcttgtctcttaaggagagac 120
Dd	120
Oy	187 attctacggtlaccatgttggtctcttggyggaagaagctctcgaagaatgactgtcacaggc 246
Dd	246
Oy	121 gtcttcogtatcatgtgtgtctcgtgaggaagaacctctgaagaatgactgcacatggc 180
Dd	180
Oy	247 atgcacggagagacatgtctggggtctctctactactctgttcctcttcttcgcgaatgg 306
Dd	306
Oy	181 atgcacgtcggagcccatgtcgtgggttctctctacttctcttagctcttcttcaccaatggc 240
Dd	240
Oy	307 cagcgttatgtataaigaagctccaataactataccaactatggtggcactgataaaggagtg 366
Dd	366
Oy	241 cagctctaatgtataaagagagctccaataaacataaccaaigttagccactgataaggagtg 300
Dd	300
Oy	367 atccatggtctcgtgaagaagattctcgaanaattcagaagaagaagatgtlgacaataatgacac 426
Dd	426
Oy	301 atccatggtccttggaagaagttctcgaatcagaagaagaacagatgtatataatgacact 360
Dd	360
Oy	427 attattgtgagagagagtggtgtgaaagtgttccacgaagaagccccctgcccactcagagaca 486

Db 361 actattatatacaggagatagtagacatgctctcagagctgacgtccgtcccatctgcgaact 420
Qy 487 aaacact---tagagagccgaggaatgcatctatctcacttactctatgvggagaaga 543
Db 421 aaactcttagtaatggaagaaggaatgcatctatactctacttacttcttggaagaaga 480
Qy 544 tccgatatcatcgtggtgcagccagccacagtgctgtgaaagaccatactacaagagctctg- 602
Db 481 acccgttctatctggtgtgcagccaaaaatgtgtgaaagaccgttcaatcagaagaatgctgt 540
Qy 603 gctgtgctctcttgcccaaatgcccgaagcctgcccggagagagttcaaaatgtgtctct 662
Db 541 gcgcgctctcttgccccaatgcccgaagcctgtccagggatgagccagatgtgtctctt 600
Qy 663 gggaacggtctctgtcttgagagcgtgtgaaatgagcactgtgacagtgccatgvcgggtctgg 722
Db 601 ggtaatgcatctgttctggatggaatgaaatgagcacaggtgtgtgtgagatgtgtggagaag 659
Qy 723 ctccaatggaacagcctgtgaaacctgacactgagggagagtaatgtatccactgagacca 782
Db 660 ctccaagccagacagcctgtgagacctgacccagggcaagtacgtacatccactgtgtacca 719
Qy 783 agcactgctctgttccatctggagagatgttaagccaagagaccttgaggagacggctctgtga 842
Db 720 agcactgtctgttccatctggagagatgttaagccaagagaccttgaggagatgctctgtga 779
Qy 843 ctgttaacgtcggctctggcggaggtgaagtgttaacatgagagatcaccaacagaaactgcaa 902
Db 780 ctgttaacgtcggctctggcggaggtgaatgttaacatgcaacccaagaaagaaactgcaaa 839
Qy 903 cgggacctgtcacacacagtgccaactgctctctgtgaatccagaacggcaaaagcctgtgcaa 962
Db 840 tggagacatgcatlacacagggccaactgctctcaacatcagaatgttaacgctttcagtcaaa 899
Qy 963 atgtcgcgcaagatctccgagggagatggaacggctctgacacagcatcaatgtcgtgtgaagc 1022
Db 900 ggtgtcagacagatctccaaggaacggaacacatctgacacagatcaatcaatgctgtgaagt 959
Qy 1023 caggcatatgagaatcttctcaacaaggccgactgtlaaanaaaccaccccggaagacgggt 1082
Db 960 caggcatatgagaatcttctcgcgcacaaaggctggaactgttaagaaacaccccggaagagagt 1019
Qy 1083 gttgtgtgtcgaagcaggtatataccgycagacggcatctgtgtccttgaatacaaccggg 1142
Db 1020 gttgacatgtgacaagcaggtctacacgggtgtgaatgtgtcgtcgtggaatacaaccggg 1079
Qy 1143 ttgtgagaaccaatgtgtgtgtgtgacagaatgtcaagatgtcacacagaacagggcccaacca 1202
Db 1080 ttgtgagaaccaatgtgtgtgtgtgacagaatgtcgaagtgtcacacagaacagggcccaacca 1139
Qy 1203 ggcctctgttaactctgtccgaagtatactgtgaaatggaatgaatgtgtgtccttatacga 1262
Db 1140 ggcctcgtgttaactctgttgcgcagcatatactgtgaaatggaatgtgtgtcgcatactcaa 1199
Qy 1263 tgtctgtcctaacgaaatgycgcgtgtgcagttccattgtgctctctcgaactaactgtgaa 1322
Db 1200 tgtctgttaactaaataatgtgcgtgtgtgaattgtgcatacttgcaacacacatgtggaa 1259
Qy 1323 agatcaaaagataatgtactgtcgaagccagactatacaggtgtgaatgtcgtctgcggag 1382
Db 1260 agtlaaagaagactgttaactgtgaagccaaactatactgtgaaatgttaactctgcgcgg 1319
Qy 1383 caggcatctacgggaggtcttcccaagaacctctgcagctcccgtaacttcttccagtctga 1442
Db 1320 caggcatctacgggaggtcttcccaagaacctctgcagctcccgtaacttcttccagtctga 1379
Qy 1443 ggaagcactgtctcgaagagctgtgtgtgaactgtgccccttccacggtgttcgcgccttggc 1502
Db 1380 ggaagcactgtctcgaagagctgt 1439
Qy 1503 tagctccttaactatgagcccggaattaaagactgtgagatcagcagggcctcattgtccaa 1562

Db 1440 tgcagccttgatgaggaagctcgggtlaaagactgvggaacaaatacggtttaatgtcccaa 1499
Qy 1563 ggttctctgtatacagctgtgtgggtcgcagacgctgcgtgttggaaacactaaagtgcac 1622
Db 1500 ggttctctgtatacagctgtgtgggtcgcagacgctgcgtgttggaaacactaaagtgcac 1559
Qy 1623 caaagtgcacagacccctcacaagagagcaggttccactctgtctctcaagacactgt 1682
Db 1560 ctcaaatgtctactctccccaagagagccaaatagtcatactccgtctctccagagccggt 1619
Qy 1683 gttcataaacaatggaagcgaaggtccctgtccagttgacatcatcaacgaacaaatgtgcata 1742
Db 1620 gtaataaacaatgaagcgaagataatcatatccagttgatacatcaagtaactaaggtgtgt 1679
Qy 1743 ccaagttatagaacaaagtgtctgtcttcccaaaaactgtctatatacccccaagaatgctt 1802
Db 1680 tcaatacatagaacaaatgtctgtcttcccaaaaactgtctatatacccccaagaacaaact 1739
Qy 1803 gggaaggtctgtgcaaaacttatacagttggcagagcaaacccaggaataacgaactgcag 1862
Db 1740 tggagaatattctgcgaataacaaatcttaacgacttggcaacaaacaatgtgtacatacaatttg 1799
Qy 1863 caagttgatacagagactcagagcttgcgtcaagtcatactatgactccataccacccagat 1922
Db 1800 caacttaatacagagactcagagcttgcgtcaagttgtcatcacccgataccatccacccagat 1859
Qy 1923 caactgtctctgtgctcaggaacaaagcccttgaaagccttggcccccaggaacagagact 1982
Db 1860 caactctctctgtgcccacagcaaacagccctccatcagctccactgtcgtgaaacaaagagact 1919
Qy 1983 cctgttcaatacgaacaaagaaggaacagcttgaaatcttaccagttgaaatcttccag 2042
Db 1920 cctgttcaacaaagaacaaagaaggaacagcttgaaaggaatcttgaatcttcaatgataag 1979
Qy 2043 aagactcgaagccttaagcttcagacctccccaaggtctgtctctcctcgtgaagaacctgtcaag 2102
Db 1980 aagactcgaagcctttagctgtgtgaatcttcccaatccactgactcgtgaaagacctgtcaag 2039
Qy 2103 ctcaagagctgaatgttgaaggtgttgaagatctgacgttgcagatccgttcttcttaacga 2162
Db 2040 ctcaagagctgaatgttgaaggtgttgaagatctgacgttgcagatccgttcttctgaatgg 2099
Qy 2163 acaaatgtatacgaatcatatacaccgaggaactctgttcttgacgtgtgggtgtgtcatatgcat 2222
Db 2100 ccaaaactgcagaatattgtgcagcggagactctgttcttgacgtgtgggtgtgtcatatgcat 2159
Qy 2223 tgaactgtactcaatgaatctccataccctgaaggtgqccgaatgtgaacactttactactctga 2282
Db 2160 tgaactgtctgtatgtatcccaaccctgggggcgctgttgaacactttactactcttctga 2219
Qy 2283 tattccggggaggttgcgggaatgtgcaatttccactcccaaatgcccactgaagaagcaagcc 2342
Db 2220 tgcctcggggaggttgcgggaatgtgcaatttccactcccaaatgcccactgaagaagcaagcc 2279
Qy 2343 aaaaggcgtgaaagaagaatgtatactacaacccgttactcttcaggaagaacgttgaagaag 2402
Db 2280 aaaaggctgtgaaagaagaatgttctctatacaaccc---tgcccttcaagaagaacacttgaaga 2336
Qy 2403 ctgcagaacactgtgtcacacgtgtgtgataccaaaccccgaaggtgtctgcattgttactcaat 2462
Db 2337 ctgcgggagcgtgtgacgctgtgtgtatlaacagatccccaaggtgtctgaaagggactacttcgg 2396
Qy 2463 gccagaactgtcagagcctgtgcctggaggagacagatatcaacggtatacaaacggggacatgg 2522
Db 2397 gccagaactgtcagagcctgtgcctggaggagacagatatcagccgtatlaaataaacggggatgtc 2456
Qy 2523 ccgcgaatctgtatacaccatgtggaacaggtgtcctatgtccacacgggtcttcaacgggaacagc 2582
Db 2457 ccttgatacgaactcctgcgcacccggagaggttaaatgtcaaacacgggtcttcaatgtggaagcgc 2516
Qy 2583 ctgcgaagcctgtcgtgcatactggaagatcttgccctgaactgttcaagcccgagctgtccga 2642
Db 2517 gttgtgaatgtgtcgtgcgggagagatttggcctgtatgtctgcacctgtgtgcctgcacga 2576

[illegible]

QY	3732	tgaqacacactcaactaattgtaacagctctccctttaaagaatcgtgtcaatgtaacctt	3782
Db	3657	cgaagacacactctgcacattgcagcatgcttttcttaatactgtaactgtgtaacgacac	3716
QY	3783	ctctagagactatgcttggaagccacactgctcattacttctagccagagacagctccaac-	3841
Db	3717	ccctgaacaagagctctggaaagcaagctgcgtccatccactgcgagccagagaccatccaac	3776
QY	3842	--aagaagccaggtcttctgtaatgaatccattctgtcagttgagacatcatgcgcga	3899
Db	3777	gaagagagacaggtctgttgaatgaagagcatctctgagtgagacatcttctcccaa	3836
QY	3900	tggatctctcatattttcttgtaacctttgagagctccctcccaagcgaacagcgtgc	3959
Db	3837	tgggtcttcatlcatlcatlctacgtcttaaaagcacccttcgccccgtgactttga	3896
QY	3960	ccactctgctctggagagatattctgtccgctgtcctgtcactgtgtgtaattgc	4019
Db	3897	ccaacctgctcttgagagagagatctctctcttgccatccctctgtgagctggcgtgtgc	3956
QY	4020	ctctgcagcttactcttactctcgcgtctaaagcagcgaacacactggttccagcgtttga	4079
Db	3957	ctgtgctgtactactctactctctgcgataaacggagaacaatcgtctccagcatltga	4016
QY	4080	-ccagaagaagacattgtatgtctgtgcttcttggaagcagacagcccaagaatctgcaa	4138
Db	4017	gtcggaaagagagacataatgtltgcagctctctggcaagcagagcctgagaaatctcgaa	4076
QY	4139	ccctctgtatagacactctacgcgcgcaccccccagagctctctctgtgaccccttcaga	4198
Db	4077	ccccttgaatgagagcacacactctagctcccccagaaactctctagaccccttcacga	4136
QY	4199	cccctggagaacaggaatctctggagaagcagacccctctggggcactgc---ggtccctgaga	4255
Db	4137	ctcttaagaacgcgcagctcttgagggaacactcttgagacactgtgaggccttgaa	4196
QY	4256	tgaagagccagaagaacacacacagctcaggtctccacgttatccagccccaagctgtct	4315
Db	4197	ggaagatgcagagcatcactcactcgtccacactggtgcacatcaactgtgaattctcagcac---	4253
QY	4316	catgtatcagttgttttaagaatgaacaacatcataaagcagccatcactacaccttct	4375
Db	4254	cagctgccttlltagaagcgtlaaagctcttlltagactaagaagcactacactcactctct	4313
QY	4376	ggttaatctcggaatctgtgcagagggctaaagagcatgttccctggaatacctggygagcc	4435
Db	4314	ggctatcctgggggtgttctctgtggtggtgagagatgtgtg-ctgtgcccacccagtaaa	4372
QY	4436	tccacctcctctgaagcctat	4455
Db	4373	gctctcctctcgtcacctt	4392
RESULT 3			
ID	AAF87120	AAF87120 standard; DNA; 3625 BP.	
XX	AC	AAF87120;	
XX	DT	26-MAR-2002 (first entry)	
XX	DE	NOV9 coding sequence.	
XX	XX	NOV: cytosolic; contrareptive; antiinflammatory; immunomodulatory;	
XX	XX	cardiovascular; casein kinase II phosphorylation site; contrareption;	
XX	XX	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;	
XX	XX	epidermal growth factor; cell development; apoptosis; cell adhesion;	
XX	XX	growth migration; cell structure; motility; cancer; immune disorder;	
XX	XX	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;	
XX	XX	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;	
XX	XX	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;	
XX	XX	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.	

OS Unidentified.
XX Key Location/Qualifiers
FH CDS 387..2024
FT /*tag= a
FT /product= "NOV9"
XX
PN WO200136638-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000WO-US31543.
XX
XX 19-NOV-1999; 99US-166336P.
PR 29-NOV-1999; 99US-167785P.
PR 08-MAR-2000; 2000US-187844P.
PR 16-NOV-2000; 2000US-0715417.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Lichenstein H, Vernet C, Fernandes E;
PI WPI: 2001-648134/74.
XX
XX P-PSDB; AAB83365.
DR
XX
XX Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
XX Claim 9; Page 32-33; 141pp; English.
XX
XX This sequence encodes the NOV9 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC cytosolic; contractile; antiinflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
XX Sequence 3625 BP; 896 A; 971 C; 914 G; 844 T; 0 other;

Query Match 34.0%; Score 1599.4; DB 23; Length 3625;
Best Local Similarity 74.1%; Pred. No. 0;
Matches 2189; Conservative 0; Mismatches 641; Indels 125; Gaps 8;

OY 1508 cctcaatcatgagcccgatataagctggatcagcagcgctcatgtccaggttc 1567
DB 571 cctcatctctctgttacaggttaagactggagacaatacagttatgtcccaagttc 630

OY 1568 ttcgctatcaagctggctggctgctccagcagctgtctgttgacaacactaaagtacacaa 1627
DB 631 ttctgtaccatgtgtgctgcctgcaccacagctgtcttctgaaacactgaattgtcttcaa 690
OY 1628 gtgccacgacccctccaaaggagagcaggtttccatctctgtcttcaggaacgtgttca 1687
DB 691 atgctacttccctccaaaggagagcaggtttccatctctgtcttcaggaacgtgttca 750
OY 1688 taaacaatlgagcggaagcttcctgcagtgatcatcatcagcaccacatgggtcttca 1747
DB 751 taaataaaggctaaagatcatatccatgtatcatcaccagtaactatggatgttcatata 810
OY 1748 ttaatagaacagttgtcttctccaaaactgtcttatccaccccaagaatgtcttgga 1807
DB 811 tcatagacaattgtctatcttccaaaattgtcttatccaccccaagaacacttggaa 870
OY 1808 gggctctgcaaaatcttactaactagtgacgaacaaacacagatataccaattcagaagt 1867
DB 871 gaattctgcaaaatcttactaactagtgacgaacaaacacagatataccaattcagaact 930
OY 1868 tgatcaggagctcaggtctgtcagttcatcatcagctcatccatccacccagtcag 1927
DB 931 taatcaggagctcaggtctgtcagttcatcatcagctcatccatccacccagtcag 990
OY 1928 tctctgtgcttgcagcaaaagcccttggaagccttgcccccagcagcagacttccctgt 1987
DB 991 tctctgtgcttgcagcaaaagcccttggaagccttgcccccagcagcagacttccctgt 1050
OY 1988 tcaatcaagacaacaagaagacaagctgaagttctatccctgaagttccacgtatccgaagt 2047
DB 1051 tcaaccaagaacaacaagaagacaagctgaagttctatccctgaagttctacgtatccgaagt 1110
OY 2048 ccaaggtttagcttcaagctccacccagctgtctgttcttggaagacccctgaagctcag 2107
DB 1111 ccaaggttttagcttcaagctccacccagctgtctgttcttggaagacccctgaaggttcc 1170
OY 2108 agctcagctgtgaagttgtggaactgtgacgtgacacacgtgtgactcttctaagaacaaa 2167
DB 1171 agctcagctgtgaagttgtggaactgtgacgtgacacacgtgtgactcttctaagaacaaa 1230
OY 2168 tgtcgaatltcaatcacccgggagctctgtttgaagtggtgtgtgctatgtcagttact 2227
DB 1231 ccttcgaatltgtcagcgggagctctgtttgaagtggtgtgtgctatgtcagttact 1290
OY 2228 gccatcatgaatccatccactagtggtgcgagtgtgacatttactaccttgcataatc 2287
DB 1291 gttctgcgattgtatccaccccttggtgggcgtgtgtgacatttactaccttgcataatc 1350
OY 2288 cggggaggtgtggaagttgtcatcttcaactcccaatgtcccaatgaagcagaagcagaag 2347
DB 1351 cggggaggtgtgtggaagttgtcatcttcaactcccaatgtcccaatgaagcagaagcagaag 1410
OY 2348 gctgtgaagaagaagttgatcatcaaacccgttacccttccaggaagacgttgaagctgtcc 2407
DB 1411 gtgtgaagcagaagttgtcatcaaacccgttacccttccaggaagacgttgaagctgtcc 1467
OY 2408 agaaccctgtgacccgtgtgtatccaaaccccccaggtgtcgtcagatgttacttactgcag 2467
DB 1468 gggagcgtgtgacccgtgtgtatccaaaccccccaggtgtcgtcagatgttacttactgcag 1527
OY 2468 actgtcagcctgtccctgggagcagatatacacccgtgttaacaacccgggcatgtgcggcg 2527
DB 1528 actgtcagcctgtccctgggagcagatatacacccgtgttaacaacccgggcatgtgcggcg 1583
OY 2528 atctgtatcaacccatgtgagcagttgcttatgtccacacccggttcaagcagcagctgtcg 2587
DB 1584 atcagtaactgtgccaacccgagagtgtaatgtcaaacaccggtcttaatgtgagcagcgtgtg 1643
OY 2588 agctcgtctgtcagatgtgagatgttggtcctgacgtcagcagcccgacgtgtcagagcatg 2647
DB 1644 agatgtgtctgtgcggtgagatgttggtcctgacgtcagcagcccgacgtgtcagagcatg 1703
OY 2648 gacagtgatgtgaggtgagatcaacaggtcccggtgaggtgctctgtgaaacaggtgtgacag 2707

[illegible]

Db	2677	accacctcgcgaatgltcaagcatgttttttctaacaatgaaccctcttgcaatgacacacccttcg	2736
QY	3788	ggactatgctcgtggaaagccaactgtcataltaactcgaagcaggaaccagctccacc---aag	3844
Db	2737	aaacgagagctcgtggaaagccaagctgtctcatcaactcgcacgagacacccaactccaacgacg	2796
QY	3845	agacacaggtctgtgtgatagtgaagatccatctctcaagctgtggagacatcatccgcgcgaatgga	3904
Db	2797	agaccaggtctgtgtgatagtgaagagccatctcgcacgttggagacatcttgcctccaatggga	2856
QY	3905	tcctccatattatctctcgaaaccttggaaactccctccccaagcaggaacagggcgcacct	3964
Db	2857	tcattcatgltcatcttccagaccctttaaagaacaccccttcgcccgctgaccttcaaccaca	2916
QY	3965	ctggcctcgtggagacagatatalactctgbcgcgtctcgtctgacctgtgttcgattgctctgg	4024
Db	2917	ctggcttggagagcagagatctctcttcttgcaatcatccctggagactggggctgttgccttgg	2976
QY	4025	cagctactcttactctccgcgttaagcagcagcagccactgtgttccagcgttttga-tcag	4083
Db	2977	ctgcttactctactcttctcgtataaacggaacaacatcgctctccagcatlttgtagtctgg	3036
QY	4084	aagaggaatctgattgctctgtgctcttggcgaagcagcagccccaagaatatcgcaacccctc	4143
Db	3037	aagaggaatctaaatgltctgacgctctcttggcaagcagcagcccggaatatctgaaaccct	3096
QY	4144	tgtatgaacctctagcgcgcgcagccccaagatctcctctgtgaacccttccacagaccctg	4203
Db	3097	tgtatgaagagacacaacccctcagctcccccagacacttccagacccttcaagactctg	3156
QY	4204	gagaaacagatctcgtgagagacagcgaacctctggggcactg---cggctctgaatggaga	4260
Db	3157	aagaacgcgcagactgtgagggcaatgaccctctgaggaacatctgtagggctcgtgacggaga	3216
QY	4261	agccagcagaagcaaacacagtcacgctctccacggtgacatccagccagcgtgtccatg	4320
Db	3217	tgccagccatcaactcaactgcacactcgtggccatcaactgtgaattctcaga---accagtt	3273
QY	4321	atcagttgtttctaaagaaatgacaacactcataaagccagacataccctaccctctgtgta	4380
Db	3274	gccttttaggaacgtlaaagtcctcttaagcactcagaagcactacatcatctctctgctg	3333
QY	4381	atctcggagatctgcgcagagcctaagggccatgtctgcctgtgataccctggggaccctcac	4440
Db	3334	atctcgggggtgttctctctctgtgggtgagagatgtctg-ctgtgcaccacccagtaacgttc	3392
QY	4441	ctccctcgaacctat 4455	
Db	3393	ctccctcgaacctt 3407	
RESULT 4			
AAAF87119			
ID	AAAF87119 standard; DNA; 2483 BP.		
XX	AAAF87119;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	NOV8 coding sequence.		
XX			
KW	NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;		
KW	cardiovascular; casein kinase II phosphorylation site; contraceptive;		
KW	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KW	epidermal growth factor; cell development; apoptosis; cell adhesion;		
KW	growth migration; cell structure; motility; cancer; immune disorder;		
KW	inflammatory disorder; cellular adhesion disorder; long QT syndrome;		
KW	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KW	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KW	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.		
XX			
OS	Unidentified.		
XX			

[illegible]

Dh	631	ttcgtaccatcgtgtcgtccgtcccaacagcttgcttcgtgaaaaactcgtaaattgatctca	690
Qy	1628	gtgcacagacccctccaaaggagacgaatttcacatctctgtctccaggaactgttca	168
Dh	691	atgctaactccctccaaaggagagccaatgatcaatccctcgtctccagagacggtata	750
Qy	1668	taaacatgagcggaaggtccgtccagtgacatcatcaaacgaacaaatggtgtatccacg	174
Dh	751	taaacataaagcctaagatcatcatccagtgatcatcatcagtaactatggtatgttcata	810
Qy	1748	ttatagaacagttgtcttctcccaaaacttgcttatccaccccaaaagttgcttggta	180
Dh	811	tcatagacaacattgtctatctcccaaaaatttggcttatcatctcccaaaagacaacttggaa	870
Qy	1808	gggtttctycaaaactctactacagtgagcagcaaaacccggatatactacaattcagtaag	186
Dh	871	gaattctgcacaaatcttaccgaacttttggccaacaaacaaatgtctatcatcaaatltagaac	930
Qy	1868	tgatcacagactccaggtctgttgcagttcaatcactgtgacttccatccaccccccactacgt	192
Dh	931	taatacagactccagttgtgtgtggttcaatcccggtcccatccacaccccgatcac	990
Qy	1928	tcttctgtccctacgagacaagaaccccttggaaagccttgcgcccaagagcaggaacttcgt	198
Dh	991	tcttctgtcccaacgagacaagaacccctccatctactccctgtctgtgacaacacagacttcgt	105
Qy	1988	tcaatccaagacaacaaagagacagcttgaaagcttaccctgaagttccacggtatccggact	204
Dh	1051	tcaacccaagacaacaaagagagacagcttgaaggagtatttgaagttccatgtgtacagagatg	111
Qy	2048	ccaaggtctttagcttccagactcccccaggtctgtcttctcttggaaagacccctgcagagctcag	210
Dh	1111	ccaaggttttagctgtgtgatacttccacatccaactgctgtgaaagacctcgtcaaggttccag	117
Qy	2108	agcttgatgtgaaagtggtgnaactgtgcagtgacatcgggtgagctcttcttaacgcgaaca	216
Dh	1171	agcttgatgtgaaagtgtgnaactgtgcagagacatcgggtgagctcttcttaagtcgcaaa	123
Qy	2168	tgtgcagatctcatcacccgggaactcgttcttgacgtgggtgtgtggtctatgtgattgact	222
Dh	1231	ctgtgaaagatgtgtgcacgctgggaagctctcttcttgacccgtgggtgtgtggtctaaagatgtgact	129
Qy	2228	gctcactcatatgacttccactcaacttggctcagatgtgtgacacttctactacactcgtatctc	228
Dh	1291	gtctgctgatactgactlcccaacccctgggggcgcgtgtgacaacttactactcttgcagct	135
Qy	2288	cgggggagtgcggaagtgtcatttcaactcccaaatgtcccatgtgaagagcaacgaag	234
Dh	1351	cgggggagtgcggaagctgtgttcaatactccagcttggccaaggtgtgagtaaaccaag	141
Qy	2348	gcgtgaaagaaagtgtatctcaaacaccggttactcttccaggaagaaacgttgaaagctgtcc	240
Dh	1411	gtgtgaaagcgaagtgtctctcaaac--tgcctctcaagaggaaccttgaaaggtgtcc	146
Qy	2408	agaaacctgtgcacgtgtgtgtatccaaacccccagtgctgtccatgttactactatccag	246
Dh	1468	ggggagcgatgtgacgctgtgtataaagatcccccaggtgtgtcgtcaagggctactctcggcag	152
Qy	2468	actgtccaggtctgcctcgtgagagaccagatacacogtgtataacaaacgggagatgtgcgcg	252
Dh	1528	actgtccaggtctgcctcgtgagagaccagatgtcccggtataaacccgggtgtgtctgtcc	158
Qy	2528	atctgtataaaccccatgtggagagtggtactatgtccacacccggcttccaaacgggaacgctgtg	258
Dh	1568	atcagcttactgcgcacacgggagagtgtaaatgtcaacaacacggcttcaatgtggagcgtgtgtg	164
Qy	2588	agctctgtctgcacgggagatttgttgaccttactacgtcagcccgcaactcctccgaagat	264
Dh	1648	agatgtgctgtgcgggagatatttgggtctatgtctgcctcgtgtgctcagaccag	170
Qy	2648	gacagtgatgtagagggatcacaggtccgggagtgctcctgttgaacaaggttgacag	270
Dh	1708	gacagtgatgtagatgtgacatcacggtcccgacgtccgtcttcttgaacaaggttgacag	176

OY	2708	ccgcttgctgtgacaaactcccaacagctgattctgcaggtgtgcaaacctgctgtctccgtgc	2767
Db	1768	gcccttcgtgtgacaactcaagcagcttctgtctgagtgtaacgcccccttcttcgcctc	1827
OY	2768	acgcacactgtacgagagaacaacacgcgtgtgtgttcttaacttgaaactaagaagtgacggga	2827
Db	1828	atgcacactgttaagagagaacaacacgcgtgtgtgtgttaaccggatttaagaagtgacggaa	1887
OY	2828	tcaacatgcacagctcgtgtgactctgtcaacaacagaacacggggcgtgtgtgcgaagtgcgtta	2887
Db	1888	tcacatgcacagctgtgtgatttctgtcaacaacagagacacggggcgtgtgtgcgaagtgtgcga	1947
OY	2888	agtgctcccaaaaagagaccccaagttctctgtgcagctgtgcaaaaagcttacaaggggatg	2947
Db	1948	gatgtcccaaaaagagagccgaaggtctccctgcagctgtgcgaagaaggtacaagaagggaag	2007
OY	2948	gctacagctgtacatagagataaacccctgtgtgacagcgtgtgtcaacgggggagatgcacatgagc	3007
Db	2008	ggcaacagctgcacacagagataagaccctcgtgtgacagcgccttaacggaggggtgtcaacgagc	2067
OY	3008	acgcacactgcagagatgcacgggcccagggcaagataagtgtgaatgtgataaagtccaactatg	3067
Db	2068	acgcacactgttaagatagatgcagagcccgggcaagcacaagtgtagtgaataaagtcacattg	2127
OY	3068	tcggggaagcggagctgtgactcgtgtagcctgtacagcagctgcgcgtctgacagcgtgtgttacaagaca	3127
Db	2128	tcgggagatggggcgtgcgaacccgtgtgagccggagcagcgtgcgcacattgacgcgtgtcttacaagaca	2187
OY	3128	acggacagctgcacacacgaatgcacagctgtgtgacagccttacttccaaagacagacgcgtag	3187
Db	2188	atgggcagtgatgcacatgcagaagcccaatgtgtgcagccttccacttcagatataccactgttg	2247
OY	3188	gaatttccatcttaagctccccacttggcgcaggtacaactgacttttgaaaagggccaag	3247
Db	2248	gggtgttccatcttaagctccccactgggcgcaggtataagctgtgacctgtgacaaagaacggag	2307
OY	3248	aagcctgtgcacaaaagaagctgcgacacatagacaccttacaacacagctctccatgtgccaga	3307
Db	2308	agcctctgtgcacaaagaagctgcgacacatagacaccttacaacacagctctccatgtgccaga	2367
OY	3308	aggccaagatatacctgtgtcgg	3331
Db	2368	agacctgtatctcttaccagaag	2391
RESULT 5			
AAAF87117			
ID	AAAF87117 standard; DNA: 2512 BP.		
XX			
AC	AAAF87117;		
XX			
DT	26-MAR-2002 (first entry)		
XX			
DE	NOV6 coding sequence.		
XX			
KM	NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;		
KM	cardiovascular; casein kinase II phosphorylation site; contraception;		
KM	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KM	epidermal growth factor; cell development; apoptosis; cell adhesion;		
KM	growth migration; cell structure; motility; cancer; immune disorder;		
KM	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;		
KM	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KM	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KM	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.		
XX			
OS	Unidentified.		
XX			
PH	Key	Location/Qualifiers	
FT	CDS	416..2425	
FT		/*tag= a	
FT		/product= "NOV6"	
XX			

XX MO200136638-A2.

XX PD 25-MAY-2001.

XX PF 17-NOV-2000; 2000OWO-US31543.

XX PR 19-NOV-1999; 99US-166336P.

XX PR 29-NOV-1999; 99US-167785P.

XX PR 08-MAR-2000; 2000OUS-187844P.

XX PR 16-NOV-2000; 2000OUS-0715417.

XX PA (CURA-) CURAGEN CORP.

XX PI Shinkens RA, Lichenstein H, Vernet C, Fernandes E;

XX DR WPI: 2001-648134/74.

XX P-PSDB; AAB83362.

XX PT Novel human polypeptides and the nucleic acids that encode them useful

XX PT for preventing, diagnosing and treating e.g. cancer, inflammation and

XX PT immune disorders -

XX PS

XX PS Claim 9; Page 24-25; 141p; English.

CC This sequence encodes the NOV6 protein. The invention relates to

CC the NOV1-NOV6 proteins, and their coding sequences. The proteins have

CC cytosolic, contraceptive; antiinflammatory; immunomodulatory; and

CC cardiovascular activities. The sequences may be used in the prevention,

CC diagnosis and treatment of diseases associated with inappropriate NOVX

CC expression. They may be used to treat disorders associated with decreased

CC expression by rectifying mutations or deletions in a patient's genome

CC that affect the activity of protein by expressing inactive proteins or to

CC supplement the patient's own production of protein. They are used to

CC produce NOVX proteins, by inserting the nucleic acid into a cell and

CC culturing it to express the protein. The DNA may be used as DNA probes in

CC assays to detect and quantitate the presence of similar DNAs in samples,

CC and which patients may need restorative therapy. The NOVX protein may

CC also be used as antigens in the production of antibodies (Abs) against

CC NOVX and in assays to identify modulators of NOVX expression and

CC activity. The anti-NOVX Abs and antagonists are used to down regulate

CC expression and activity. The anti-NOVX Abs are used for detecting the

CC presence of NOVX in samples. Disorders that may be prevented, diagnosed

CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,

CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites

CC characteristic of serine/threonine kinases, and are used to treat

CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular

CC proliferation and corticoreception). NOV2-3, NOV6 and NOV8 are homologous to

CC the epidermal growth factor (EGF)-like super family and are involved in,

CC e.g. regulation of cell development, apoptosis, cell adhesion, growth

CC migration, cell structure and motility and protein management, and are

CC used to treat cancers, inflammatory disorders, immune disorders and

CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like

CC fibillin proteins and are used to treat cardiovascular disease e.g.

CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.

XX

SQ Sequence 2512 BP; 623 A; 657 C; 651 G; 581 T; 0 other;

Query Match 26.4%; Score 1243.2; DB 23; Length 2512;

Best Local Similarity 80.5%; Pred. No. 0;

Matches 1468; Conservative 0; Mismatches 353; Indels 3; Gaps 1

Db 1508 cctcaatcatgagcccgagatgaagactggatcagcagggcctcatgtccaggttc 1567

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

Db 600 cctctgatactctgtctacaggttaagaagcctgggacaataagtgtaatgcccaagttc 659

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 1568 ttccgtcatcaagctgtgtggcgtccagcagcgtctgtgttgacaacactaaagtgcacaca 1627

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

Db 660 ttccgtcatcaatgtgtgcgtccagcagcagctgtcttcggaaaacctgaattgtatccaa 719

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

QY 1628 gtgcacagaccctccaagagagagccagttccatctctgtctctccaagagacactgttca 1687

111 111 111 111 111 111 111 111 111 111 111 111 111 111 111

Db 720 atgtctactctccctccaagagagacaaatagctacatctccgtctcccaagacagctgtata 779

QY 1688 taacaatgaagcgaaggtccctgctccagtgacatcatccagcacaatggcgtcatccag 1747
 Db 780 taacaataaagctaagatcatatcatcagtgatcatcatcagtaactatgagatggttcata 839
 QY 1748 ttatagacaagtgtgctctcccaaaacttgcttatcaaccccccaagaatgacctgggca 1807
 Db 840 tcatagacaatgtctatctcccaaaatttgcttatcatctcccaagaacaactctggaa 899
 QY 1808 ggggtctccaaatcttctactaagtgagcgaacaacacagatatacccaattccaaagt 1867
 Db 900 gaattctgcaaatcttctcagactcttgcaacaacaatggtctatacatcaatttgaact 959
 QY 1868 tgatacagagactcaggtctgctgctcatcatcagatccatccacacccagagctg 1927
 Db 960 taatacagagactcaggttctgctgagtgctcatcaccgcatcccatccacccagctac 1019
 QY 1928 tctctgacctacggaacaagacctggaagaccttgccccagagcagcagactctctgt 1987
 Db 1020 tctctgccccacccgacccaagccctccatctgcccctactgctgtaacaacagactctctgt 1079
 QY 1988 tcaatcaagacaacaagacaagaactgaagcttactgaagttccacgttgatccgagact 2047
 Db 1080 tcaacacaagacaacaagacaagaactgaagagatattgaagtttcaatgtagacgagatg 1139
 QY 2048 ccaaggtcttagactcagacacctccccaagtgctgcttccctggaagacccctgcaagctcag 2107
 Db 1140 ccaaggtcttagactgtagtcttcccatatccactgctggaagacccctgcaagttcag 1199
 QY 2108 agctgaagtgtgaagtgtagaacctggcagtgacatcgtgtgaagcttcttctaaacgacaaa 2167
 Db 1200 agctgaagtgtgaagtgtgaagctggcagagacatcgtgtgaagcttcttctgaatggtc 1259
 QY 2168 tgtgcaggttcatcacacgggagactctgtttgaagtggtgtgtgctatgagcatgact 2227
 Db 1260 ccgtgcaagaatctgtagcagcgagacctgttctgagacctgggtgtgcttcaagcatgact 1319
 QY 2228 gactcatatgaatctctactcactgaagtgagcagatgtagacacttcttactccctcgtatlc 2287
 Db 1320 gctcgtgattgattccaccccggtggggcgctgtgtagacaccttactcttcgtgctc 1379
 QY 2288 cgggggagtgcggaagtgatcttctactcccaaatgcccactgaagagcgaagcacaagg 2347
 Db 1380 cgggggagtgcggaagtgatcttctactccacagctcgaagtgaggttaaaccaaaagg 1439
 QY 2348 gctgtgaagaagaagtgatcttctactaacccgcttactcttcaagagagagctggaagctg 2407
 Db 1440 gctgtgaagaagaagtgatcttctactaac--tgcccttcaagaagaaactggaagctg 1496
 QY 2408 agaactgtgacacgtgtgatacccaaccccaagtgctgcatatgattactcatgacag 2467
 Db 1497 ggggagcgtgtgacacgtgtgatacccaagctcgaagtgctgcaagggctactctggcgag 1556
 QY 2468 actgtcagacctgcccctggagagacagatatcacggtgtaaacacccgggcatgtgcgag 2527
 Db 1557 actgtcagacctgcccctggagagacagatgtccccgtgtaataaacgggggtgctccttg 1616
 QY 2528 atcgttacaacccatgggaagtgacctatgcccacacccggtcttcaacgggtagacgtctg 2587
 Db 1617 atcgttactcggcacccggagagtgtaaatgcaacacccggtcttcaatggagcggtgtg 1676
 QY 2588 agctactgtgacacgtgtgagacttgagcctgactgcaaccccgacgtcagcgacgtg 2647
 Db 1677 agatgtgctgctggcgggagatttgagcctgactgtctgctcctgtgctgctcagaccag 1736
 QY 2648 gacagtgatgtagagggatcacaagctccggggaggtgacctgtgtaaacaggtgagacag 2707
 Db 1737 gacagtgatgtagatgtagcaccaggtccgggcagtgacctgtgtaaacaggtgagacag 1796
 QY 2708 ccgctctgtgtagacacatcccaacagctgattcagagtgtagacacgtgctgtccgagc 2767
 Db 1797 gccctctgtgtagacacatcccaacaggttctgtcagtgtagacacgtcctctgtctgtc 1856

QY 2768 acgccaactgtacgagacaacacacgtgtgtgtgtaacttgaactacgaaggtagcggga 2827
 Db 1857 atgccacacctgtataggaacaacacacgtgtgtgtgtaacttgaactacgaaggtagcggga 1916
 QY 2828 tcaactgtcacagctgtggaactctgtcacaacagacaacacgggggctgtgcgaagctgtcta 2887
 Db 1917 tcaactgtcacagctgtgtgttctgtcacaacagacaacacgggggctgtgcgaagctgtcca 1976
 QY 2888 agtgtctccagaagaagcaccacagctctctgtgagctgtgcaagaagaaggtctacaagggtgag 2947
 Db 1977 gatgtctccagaagaaggtgtctctgtcagctgtgccaagaaggatatacaagggtgag 2036
 QY 2948 gctacagctgtcatagatatagacccctgtgtcagacggtgttcaacggggatgtccatgagc 3007
 Db 2037 ggcacagctgtcacagaagatagacccctgtgtcagacggtcttcaacggggatgttccagagc 2096
 QY 3008 acgccaactgtcacagatgacgggcccaggaagacataagtgtaagttaaaagtcacatag 3067
 Db 2097 acgccaactgttaagatgacaggtcccggtgaagcacagaatgtgtaaaagtcacatag 2156
 QY 3068 tcggggagcggagtgtagactgtgtgagcctgtgagcagctgtccgctgcagcgtgtgttaccagaca 3127
 Db 2157 tcggagatgtggtgtgaactgtgagccggagcagctgtcccatltgaccgctgtcttaccagaca 2216
 QY 3128 acggaacgtgtgccaccacagatgtcagcgtgtgcagaccttacttccacagacacgacgttag 3187
 Db 2217 atgggcagtgtagcagatgtcagacgtccaaatgtgtacaccttccatccagatlaaccactgtg 2276
 QY 3188 gaggatctcatctacgtctcccatctgggcagtgacataacactgtgacatttgcacaagccaaag 3247
 Db 2277 ggggtgttccatctacgtctcccatctgggcagtgacataacactgtgacatttgcacaagccaaag 2336
 QY 3248 aagcctgtgtccaaagaagctgtgacacatagcaccacttacaacacagctcttctatgtccaga 3307
 Db 2337 aggcctgtgtccaaagaagctgtgacacatgtgacacatcttaccataacacagctcttctatgtccaga 2396
 QY 3308 aggcgaagtatcaccctgtgtcgg 3331
 Db 2397 agacctgtatctccttaccaaag 2420

RESULT 6
 AAF87113
 ID AAF87113 standard; DNA: 2011 BP.
 XX
 AC AAF87113:
 DT 26-MAR-2002 (first entry)
 XX
 DE NOV2 coding sequence.
 XX
 KW NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
 KW cardiovascular; casein kinase II phosphorylation site; contraception;
 KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
 KW epidermal growth factor; cell development; apoptosis; cell adhesion;
 KW growth migration; cell structure; motility; cancer; immune disorder;
 KW inflammatory disorder; cellular adhesion disorder; long-OR syndrome;
 KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
 KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
 KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
 XX
 OS unidentified.
 XX
 FH key Location/Qualifiers
 FT CDS 877..1824
 FT /tag= a
 FT /product= "NOV2"
 XX
 PN WO200136638-A2.
 XX 25-MAY-2001.
 PD 17-NOV-2000; 2000WO-US31543.
 XX

Db 1475 ggggtgtccatctacgtctccaccatggtggccagctataagcttgacctcttgacaagccagag 1534
Oy 3248 aagcctgtgccaagaagctgtgacattagaccattacaacagctctctctgtgccaga 3307
Db 1535 aggcctgtgccaagaagctgtgacccaaggacaacctacaacagcctctctatgtccaga 1594
Oy 3308 aggcacaatatacctgtgctcgccgagctgtgctgagagatggtggcggtgtgctctaccga 3367
Db 1595 aggcacaagtaccactgtgctcagcaggtgtgctgagagaccgggggtgtgctctaccga 1654
Oy 3368 ctatgtatgctctcagaagtgtggtgcacaacgtgtgtggagctgtagactacgattcca 3427
Db 1655 cagcctgtgctccaccagactgtgctgtgtgtgtgagtagtgactatgaccca 1714
Oy 3428 gggccaacaagaatgtaattgtggatgtctctgttaccggatggaatgtaagtgaactgca 3487
Db 1715 gaccacaacaagaatgtaattgtggatgtctctgttaccggatggaatgtaagtgaag----- 1762
Oy 3488 cctgcaagcgagctatgtgggagatggtctctcgttgagtggaacctgtcgcagtlcc 3547
Db 1763 ----- 1762
Oy 3548 tcatgtctctccctcgtctacacaactctcgcagagaagtgtgctgtcttcttccaagagct 3607
Db 1763 ----- gaagtgtcgtgcattcttccaacagct 1787
Oy 3608 cagcccgagcagcagcatttctgaaacacgtgacctgacctgcctcgtcgtgacacctgt 3667
Db 1788 cagctcgagcgcgtgcatcttccagaacacctgagacctgtccatccgcggacacctct 1847
Oy 3668 ttgtgccacaagaacagtgtggtctaccggaaataaagacctgtcgtcgccggagcattgagc 3727
Db 1848 ttgtgccacagaaacagtgtggtcgggaggaatgtagacctgtgtcgtggcgagcctcgagc 1907
Oy 3728 accactctacctaigtctcaagctctccttcttacaatgacctgtgtaagtcttcttctga 3787
Db 1908 accactctgccaatgtcagcatgttcttctacaatgacctgtgtaagtcttcttcaatgca 1967
Oy 3788 ggaactctgtggaagcacaactgctcattcattcagccagagac 3831
Db 1968 aaacgagcgtgggaagcagaagctgtcattcattcagccagagac 2011
RESULT 7
AAF87114
ID AAF87114 standard; DNA; 1804 BP.
AC AAF87114:
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX NOV3 coding sequence.
DE
XX
XX NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KM cardiovacular; casein kinase II phosphorylation site; contraception;
KM serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KM epidermal growth factor; cell development; apoptosis; cell adhesion;
KM growth migration; cell structure; motility; cancer; immune disorder;
KM inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KM cardiovacular disease; hypertrophic cardiomyopathy; marfan syndrome;
KM therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
XX NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.
OS
XX
XX Unidentified.
FH
FT
FT Key Location/Qualifiers
CDS 877..1611
FT /*tag= a
FT /product= "NOV3"
XX
XX
XX W0200136638-A2.

PD 25-MAY-2001.
XX
XX 17-NOV-2000; 2000WO-US31543.
PR
XX 19-NOV-1999; 990S-166336P.
PR 29-NOV-1999; 990S-167785P.
PR 08-MAR-2000; 2000US-187844P.
PR 16-NOV-2000; 2000US-0715417.
XX
XX (CJNA-) CUBAGEN CORP.
XX
XX Shinkels RA, Lichenstein H, Vernet C, Fernandes E;
PI WPI: 2001-648134/74.
DR P-PSDB: AAB93359.
XX
XX Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
XX Claim 9; Page 14-17; 141pp; English.
XX
XX This sequence encodes the NOV3 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC Cytostatic; contraceptive; antiinflammatory; immunomodulatory; and
CC cardiovacular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g., regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovacular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
XX
XX Sequence 1804 BP; 426 A; 487 C; 521 G; 370 T; 0 other:
SO
Query Match 21.6%; Score 1016.4; DB 23; Length 1804;
Best Local Similarity 82.0%; Pred. No. 8.1e-278;
Matches 1183; Conservative 0; Mismatches 256; Indels 3; Gaps 1;
Oy 1868 tgatacagactcagagctgtgctgcatcattcagactcaccacccagctcactg 1927
Db 158 tgttcagactcagagctgtgctgagtgcatcattcaccatccacacccagctcactg 217
Oy 1928 tcttctgctcagagcagaagcctgtggaagctgtgccccagagcagaagcattcctgt 1987
Db 218 tcttctgcccacagcagcaagccctccatgacctcgtgaaacaagagctcctgt 277
Oy 1988 tcaatcagaacaagaagcagctgtaagcttactactaaagtctcagctgtagcagact 2047
Db 278 tcaaccaagacaacaagaagcagctgtaagcttactactaaagtctgtaagcagatg 337
Oy 2048 ccaagctttagcttcaagactctcccaagctgtcctctcgtggaagacctgcaagctcag 2107


```

Db      338 ccaagttttagctgtgtgacgtcttcaccaatccacatgctcgtggaagaccctgcaagtttcag 397
Oy      2108 agcttgatgtgagtggtggaactgcagtcgtacatcgctgagctcttcttaagaagcaaa 2167
Db      398 agcttgatgtgagtggtggaactgcagtcgtacatcgctgagctcttcttaagaagcaaa 457
Oy      2168 tgtcagatcatacacccggggaactctgttctgacgtgggtgtgacctatgcatgact 2227
Db      458 cctgcagaattgtgcagccggagactctgttctgacctgggtgtgacctatgcatgact 517
Oy      2228 gactactcaagaattctaccatggtgtgacctatgcatgactcttactactctgatatc 2287
Db      518 gctctgcatgatttccacacccctggggccgctgtgacacattactactctgatatc 577
Oy      2288 cggggagatgcggaagtgtgacttcttcacatcccaatgccacatgaaagcaagcaag 2347
Db      578 cggggagatgcggaagtgtgacttcttcacatcccaatgccacatgaaagcaagcaag 637
Oy      2348 gctggaagaagaagtgtatcatcacaccgttacccttcagaaggaacgtgtgaaggtgcc 2407
Db      638 gctggaagaagaagtgtatcatcacaccgttacccttcagaaggaacgtgtgaaggtgcc 694
Oy      2408 agaactctgcacacgtgtgataccaacccccaggtctgcacatggttactatcagcag 2467
Db      695 gggagagatgcagctgtgtatatacagatccccaggtctgcagaaggttactctgcggcag 754
Oy      2468 acgtgcaggtctgcggcggaagacagatatacaccgtgttaacaacggggcgtgcggcg 2527
Db      755 acgtgcaggtctgcggcggaagacagatatacaccgtgttaacaacggggcgtgcggcg 814
Oy      2528 atctgtacacacccatggaagcagtgctatgtccacacccggttcacagggaacagctgcg 2587
Db      815 atcagtaattctgcggcagcggagagtgtaatgtacaacccggttcacagggaacagctgcg 874
Oy      2588 agctctctgtgcagtggaagattgtgacctgtacgtcccgacgtctgcggagatg 2647
Db      875 agatgtctgtgcggcggaagattgtgacctgtacgtctgcggagcagcagcaacgcg 934
Oy      2648 gacagctgtgagagggatgacacagctccggggaggtgctctgtgaacaacgggtgagacag 2707
Db      935 gacagctgtgagagggatgacacagctccggggaggtgctctgtgaacaacgggtgagacag 994
Oy      2708 ccgctctgtgtgacacatcccaacagctgatatcgcaggtgtgcacacgtctgtctcgtgc 2767
Db      995 gccctctgtgtgacacatcccaacagctgatatcgcaggtgtgtcgtctcgtctcgtgc 1054
Oy      2768 accgcaactgtgacaggaacaacacggtgtgtgatacttaactgaactacaggaagtacaggga 2827
Db      1055 atgcacactgtgaaggaacaacacggtgtgtgatacttaactgaactacaggaagtacaggga 1114
Oy      2828 tcacatgacacagctgtgactctgtcaacaacaggaacgggggtgtgcagaggtcgta 2887
Db      1115 tcacatgacacagctgtgactctgtcaacaacaggaacgggggtgtgcagaggtcgta 1174
Oy      2888 agtgcctccagaagaagcaacacagctctgtcagctgtcaagaagaaggtcacaaggggatg 2947
Db      1175 gatgcctccagaagaagcaacacagctctgtcagctgtcagaagaaggatcacaaggggatg 1234
Oy      2948 gctacagctgtatagagataacccctgtgcagagcgtgtcaacgggggtgagtcagagc 3007
Db      1235 ggcacaagctgacagagataaacccctgtgcagagcgtgtcaacgggggtgagtcagagc 1294
Oy      3008 accgcaactgtgacagagataaacccctgtgcagagcgtgtcaacgggggtgagtcagagc 3067
Db      1295 accgcaactgtgacagagataaacccctgtgcagagcgtgtcaacgggggtgagtcagagc 1354
Oy      3068 tcggggagcaggtgtgactgtgagcctgtgacagcgtgcgcgtcgcacgctgtgttacaagaca 3127
Db      1355 tcggggagcaggtgtgactgtgagcctgtgacagcgtgcgcgtcgcacgctgtgttacaagaca 1414
Oy      3128 acggaacatgtcacccaatgtccagctgtgcagaccttacttccaggaacagaccgttag 3187

```

```

Db      1415 atggcagctgcacatgcagacgccaatgtgtgacctccactccagatataccactgtg 1474
Oy      3188 gagtattccatcactagctcccaactgtggcaggtacataacatgcacattgtgaagaagcaag 3247
Db      1475 ggggtgtctacatcactcccaactgtggcaggtacataacatgcacattgtgaagaagcaag 1534
Oy      3248 aagcctgtgcagaagaagctgcagaccaatgcacacacacacacacacacacacacacacac 3307
Db      1535 aagcctgtgcagaagaagctgcagaccaatgcacacacacacacacacacacacacacacac 1594
Oy      3308 ag 3309
Db      1595 ag 1596

RESULT      8
AAA57365
ID      AAA57365 standard; DNA; 1259 BP.
XX
AC      AAA57365;
XX
DT      03-OCT-2000 (first entry)
XX
DE      DNA encoding a human hyaluronan-binding protein, designated BM-HABP.
XX
KW      Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;
KW      Proliferative condition; metastasis; inflammation; ischemia;
KW      host defence dysfunction; immune surveillance dysfunction; arthritis;
KW      multiple sclerosis; autoimmunity; immune dysfunction; allergy; ss.
XX
OS      Homo sapiens.
XX
FH      Key
FH      Location/Qualifiers
FT      CDS
FT      199..1257
FT      /*tag= a
FT      /product= "hyaluronan-binding protein"
FT      /transl_except= (pos: 478..480, aa: Xaa)
FT      /transl_except= (pos: 667..669, aa: Xaa)
FT      /transl_except= (pos: 1105..1107, aa: Xaa)
FT      /transl_except= (pos: 1138..1140, aa: Xaa)
FT      /transl_except= (pos: 1156..1158, aa: Xaa)
FT      /transl_except= (pos: 1168..1170, aa: Xaa)
FT      /transl_except= (pos: 1171..1173, aa: Xaa)
FT      /transl_except= (pos: 1249..1251, aa: Xaa)
FT      /note= "Xaa is an unspecified amino acid; no
FT      termination codon given"
XX
XX
XX      MO200039166-A1.
XX
XX      06-JUL-2000.
XX
XX      20-DEC-1999; 99WO-US30462.
XX
XX      23-DEC-1998; 98US-0113871.
XX
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      (AMNA-) AMERICAN NAT RED CROSS.
XX
XX      Hastings GA, Liau G, Tsifrina E;
XX
XX      WPI: 2000-452376/39.
XX      DR
XX      P-PSDB: AMY93913.
XX
XX      New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP,
XX      OE-HABP and BM-HABP, useful for treating proliferative conditions,
XX      metastasis, inflammation, ischemia, arthritis and multiple sclerosis -
XX
XX      Claim 2; Fig 4A-B; 457pp; English.
XX
XX      The present sequence encodes a hyaluronan-binding protein. The
XX      specification describes four hyaluronan-binding proteins, known as
XX      WF-HABP, WF-HABP, OE-HABP, and BM-HABP. The polypeptides are useful
XX      for treating diseases such as proliferative conditions, metastasis,

```


QY	25	ctggagcagatgcccgactattccatttccggagttaactatttcaatcaacccctggca	84
Db	2333	ctggacttggcttccttccttcacgaccttcctgggaattgtctgaagacaaccttggcttgg	2392
QY	85	atgagcaatcgagctctcagatgtcttaactgtgtcttcgtgcacaacaatgaaagcatcgaa	144
Db	2393	ccccgaattggaagctctccactgtcctcaaccatttcttggccccaacccgtctcccgga	2452
QY	145	aactatatacggagagaagaagccacatcctctaaaggaagatactctaggttaacatgtg	204
Db	2453	gcc-----cagggcaacagcagtcacctctgacgcagacacagctgctggccacatctg	2503
QY	205	gtctctgggggaaaaagctcttcgaagaatgtcttcataacgctctgacacgcgaagagacatg	264
Db	2504	gtctctggggggaagccctcttcatactgaaacccctcgggaaggctgaaacccgaacaccttc	2563
QY	265	ctcgggtctctccacaccttccttcctcttccttcctccgcacatgacagctgtatgttaatgaa	324
Db	2564	ctgggcccctggcccaactggtatcgctctcttaaacacacagctggtccagcctctggagtgaaacat	2633
QY	325	gtctccataaactacacccaatgtgtgcaccatgataaaggagtgatccaatgtctcggagaaa	384
Db	2624	gtgcacactggaagccccaatgtcggagggccctctggccctgcatgtgtctgtctgggg	2683
QY	385	gttcttggaatttaagaagaacagatgtgaaataatgaacccaattattgtagaaggaggag	444
Db	2684	gtctcttgcagctgggtctcaagtctgtctgtctgtacatgaacacgctgaggccctcgggggaanaa	2743
QY	445	tgtcggaagaagtctcccaacgaagcccccctgcaccactcgagacaaacccaacttaagaagagcg	504
Db	2744	tgtgttaaacctgcacccaagagatcttcgctgtacatcagggtctccaagctctgcagagacacacc	2803
QY	505	aggaaatgcatactcatccatctactctacatgagggagaagagatccgtattcatcgggtgtccag	564
Db	2804	aggaaagaagctgtgtctacccaatctgtgctt-----ctcctctcccggggtgtgctct	2854
QY	565	ccacagtggtgtgggaacatcatcttaacaagagccgtctgt-gtctgctctctctggccaca	623
Db	2855	tacacatgtgtccagaagaagatccacaggtgtccggaactgtccgcccctgtgtctcttgggaagctgt	2914
QY	624	tgcgaagctgtcccgccggagagatgcataaagtgtgtctctgtggaaagggctctctgtctgga	683
Db	2915	tgtgagagcattgcacgaaggggtctaaagggggtgtgtgtccagggccaatgggcaggtgcc-agg	2973
QY	684	cgggtgtgaatctgcaactgtgcaatgtgcaggtgtccaggtgtggcttcaaattggagacacgttga	743
Db	2974	caggttctctgggcagcggggaggtgtccactgtccacaggaaggcttccatagtgaacagcgtctga	3033
QY	744	aaacctgacatcgagggggaagatagtataccagctgtgacacgaacagatcgtctgtgtccatgg	803
Db	3034	ggtgtgtgtgagcttggccgtctacgggtcccaactgtcacccggaggtgtgtgtatctgtgtccatgtg	3093
QY	804	gagatgtagaccagaagacccctctgggagagacggctctctgtgactgtgtacgtctggctgtgcagg	863
Db	3094	gtctgtgcagaaggagggtgtctgcaaggggagcgaagactgtgtctgttaacgttggctgtgcagggt	3153
QY	864	agtgaaggtgtgacatctgagatcacacacagaacaactgtgacacgggacatctgtcaacacagtg	923
Db	3154	ctctccgctgtgacacagaanaatcaccaagccctcagtgctccctaggaagtgtgcagccccaatgc	3213

OY	924	caatctgctcttgatccagacgagcaaacgctccgagcaaatgctgcgacagattccgagg	983
Db	3214	caactctgctgacgaactcgccggaacccctccacccctgcgcgcctgctcgagataatccgg	3273
OY	984	gaatcgacaagctctgcacagcccatcatgctccgtgtagaacacagcaatggagaatttctac	1043
Db	3274	caatcgacattctctgttcagagaggttgaaacccctgcgcgcacagccatggggctctcc	3333
OY	1044	aaagggccgactgtaaagaacccacccaggaaccgggtgtgtgtgtgcaagagacta	1103
Db	3334	tcatcgcaactatgttaacaaagltgacaccttcggagagagacatgacacctgcgcgagtctta	3393
OY	1104	taccggagacgacatcgtgtgcctttaaatacgaaccggtgtgttgagaacatagtatgctc	1163
Db	3394	catgtgcgacggagagactgtgcgcaggaataatacagctgtctctatccacacggggctg	3453
OY	1164	tgcacagaatgtcagatgcaacagacagagcccaacagacgctctgttaactgtgc	1223
Db	3454	ccaacttcaacgcgagatgcatctccccccttcggccccaagagatctctctgcagctgcgtga	3513
OY	1224	gaagtatacctcgagagatggaa---aggctgcgcctgtatcaatgtctgtccctaaagaa	1280
Db	3514	gggttaacgcgggagatgagcatccggacccctgcgagctccttgaaacccctgtcttaagaaca	3573
OY	1281	tgcgcgctgcagatccatttgcctctctgtacaactacatgacagaataatcaagatatgtac	1340
Db	3574	tggagagatgacagcccatatgtccacctgtcaacaagacaaaggatgagacagagatatc	3633
OY	1341	ctgcagaacgcagactaac---gggtgatggaatgctgttcgcggggacgaactatccggga	1397
Db	3634	ctgcgataacgcgcacacacgcgtgggggagagccctcacatgcgtgcgcgagctgcgcgga	3693
OY	1398	gcttcccaagaaccccttcagacccctccagacttcttccagatgtgcagaggaatgtgtcgc	1457
Db	3694	gtcccttgaaggataagatatgctcattcttcttcaagccctccctcctggatat-----aa	3747
OY	1458	agagctgtgtgacccctgcgcctccctaacgcgtgtgtgcgccttgtcttagctccctcatca	1517
Db	3748	ggaactcacaaggcgatgagcttcttccacactctctgtgcgcgcagacgaatactaatgacaa	3807
OY	1518	tgaagcccgagatlaaagaactggaatacgaacagagccctcaatgctccagattctcctaca	1577
Db	3808	ccctgtccagagatgagctgtgcgcggatctgtgtgcatactgcagcgcggttctgctaca	3867
OY	1578	cgatgtgggctgcccagcagctgtctgttggacaacctaaagtgcacacaagtgcacagac	1637
Db	3868	cgatgtgtgcctgcgcgtctgcgcgcgcgagacgaactgcctgcgcgcaggggtacgcagcgc	3927
OY	1638	cctccaagagagaacagtttccatctctctctctcaagagacatggtgttcatlaaacaatga	1697
Db	3928	cctcccaaggacacccacacgcgtcttccagcgagagggagggacgacatactccatagact	3987
OY	1698	ggcgagaagcttccctgcacgtatcacatlcagacacaaatggcgtlcatcaacgtlatagaaa	1757
Db	3988	cgcgcgcgctgtgtgagcagcgacatagagcgcggtgaacgcgcatcctgcacatltgacgcg	4047
OY	1758	gttgcgtctctccaaaacttgtcttatacacccccacaaagtgtccttgggcaggtttctgca	1817
Db	4048	tgtctctgtgtcccccgcgaagcgctgcacatctggagccctgtatgtctccatctccgagag	4107
OY	1818	aaactttactacgttgcacgaacacacgcgatatatacnaattcagaagaattgtatcacgaa	1877
Db	4108	aaatgtcacgcgcgcgcgcacagagcttctgtttacaagatcttccagcgctctccctgaagtc	4167
OY	1878	ctcagactgtgtcatgtatcaatcatcgtatccatccacacccagatcaatgtctctctgc	1937
Db	4168	ggccgcgcctctgcgcctgtcttgagagagatcccatagcccttcaacaatgtctgtgcc	4227
OY	1938	tacggaacaaaccccttgaaagccttgcgcccaagacgcagagacctcctgttcaatcaaga	1997
Db	4228	cacagaacgcgccttctcgaactgttcttccctccgaatgcgcagcgcctgtgttatacagaga	4287
OY	1998	caacaagaacagcttgaatcttactccctgaagtttccacgtgataccgagatcccaagctt	2057

Db 4288 ccacccgtgacaagctagaacacatctctgcygagccacatgatctgcagatgtccagagcctt 4347
OY 2058 agcttaagacctcccccaggtcgtctctctcttggaagacctctcaagagctcagaagcgtatgt 2117
Db 4348 ggcatactgacactgcacacaccttgagccaccttcgaacatcatgtagagacccccacctctct 4407
OY 2118 gaagttgtgaacttgcagatgacatcgtatgagctctctcttcaaacgaacaatgtgcaat 2177
Db 4408 ---ctctctgcaagccgaacccgagcccgctgagctcatagtgtgtagagatgatagtctcgcat 4464
OY 2178 cataccagcgagacctctgttttgaactgtgggtgtgtgacctatggaacttgaactgtactcat 2237
Db 4465 tgtgtcaagcgacacttgacctttgaggggtgtgacctatgtgcatgacacagcagctgtcgtga 4524
OY 2238 gaatccatccactagttgtgcagatgttgacacttttaactacctctgcatatctccggggaggtg 2297
Db 4525 gccacactgacctgtgtgtctgcgtctgtgacacttgagacacccggccctgcgactggaacac 4584
OY 2298 cgtgaagttgcatl-----ttaaactccaatgtccacatggaagagaagcc 2342
Db 4585 ctgcagacatctgtgtgggtctgtagcaccacctgtctcgtgaaggtgtcaagtagcagagcagccc 4644
OY 2343 aaagggc-----gtg 2352
Db 4645 tgaagcctctgctgagcttctlaaccgaaagtctctgagcgtccctctcgtcgtacatctctggg 4704
OY 2353 aagagaagaatgtatactcaacccgttaaccttcaaggagagac-----gtcggaag 2402
Db 4705 attacgaagcgtctctgggttccaccccaagcctlttggttagagcccaagcgtcgtgacaggtg 4764
OY 2403 ctgcagaacacctgtgcacccgtgtgtgtatccaaacccagaggtgcgtcatgttactcat 2462
Db 4765 ctgcagacccgaatgtgttcaacacacacttgaaagcccaagcgtcgtccctgtgtcatatgtg 4824
OY 2463 gccacagctgtcaagcgtgcctctggagagccaagatacacacgttgttaacaacccggagcatgtg 2522
Db 4825 cagtgtgtgccaagctctgccttgcygagcccaagcagccctgtatgtagcagctgtgctgtg 4884
OY 2523 ccggagatctgtatacacacccatlgagacagtgtactatgtcacacacccgcttcaacggagacagc 2582
Db 4885 catgtgaagcagatgtatgtatgtgagcagctgtctgtctgtccgttcaagctlttgctgtggaacgc 4944
OY 2583 ctgcagagctctgtcgtgcagatgagaaatttggtgcgtgcagctgcagcccgcaagctctccga 2642
Db 4945 ctgtggaactctgtgtcctctgtgctcttggtgcccattgttaagcctgcgcgtgcactgtc 5004
OY 2643 gcatgtgaacagtgtgaagagggatcacagagctccggggaggtgcctctgttgaacaacaggtg 2702
Db 5005 gcatgtgcgcgtgtatgtatgtatgtgagccctltggggggtctgtgcctcctgtctgtgtagaaggtcgt 5064
OY 2703 gacagccgcgtctgtgtgtgacactcccaagctgtatctgcagatgtgtgacacactgtctc 2762
Db 5065 gactgtggccacgcgtgtatgtatgtcaactgtgagcgtctgcaagcctgtgtgtacccccaccctgtgc 5124
OY 2763 cgttcaagcacaacctgtacggaacaacacagctgtgtgtgttaacttgaactaagaaagtgtga 2822
Db 5125 acccgaaagcgtgtgtgcgtctgcaagcaacagctgtgtatgtcagcgtctgggtcatltgaaaggtga 5184
OY 2823 cgtgatatcaatgtgacagctcgtggaactctgcgaacaaacagaaacagcggtgctgtgcgaagt 2882
Db 5185 tggcgcgtgtgtgtacagttgtgagacactgtgtcagagacgtgtgtgtcagltgtgagca 5244
OY 2883 cgtctaaatgtccccaagaagaagcaccagaagtctctctgcaagctgcagaagaagcgtacaaggg 2942
Db 5245 cgtcaaacgttagcagaagttagaacaatagttaactgttaacctgtgcctgcgcgcagctcaggg 5304
OY 2943 ggtatgtctaaagctgtgcatagagataagacccctctgtgcagagcgtgttcaacaggtgatatcca 3002
Db 5305 tgtatgtctgtagagctgtccggggtccgcaacccctctgtcacagatgtgcacccgctgggtctgtcag 5364
OY 3003 ttagagacgcacacatgtacagatgtacgtgcccagagcaagcataagtgtgaattgtaaaagtca 3062
|||||

Db 5365 cgaagcagccaaactgtctgtgaagcaccggcctgtgaacacacagcgtcgtgtgtatgtgcacgcag 5424
OY 3063 ctatgtcggggagagtgtagctgtgagccgtgaagcagctgcgc-----tcgaacctgtct 3119
Db 5425 ctacgttaggcgatgtgactgtcagttctgtgagaggtctggaaccactctgtgtgacgcgtctct 5484
OY 3120 acaagacaacaggaacagtgtccacccagatgcccagctgtgtgaagacctactactacagagacac 3179
Db 5485 gggcagcagcacaagccctgcacactcagatgccaatgtgtacatgtaactgtcagactctccagagaa 5544
OY 3180 gaaccttagagatattcatatcagctctcccaactgtggccaggtatagaactgtacatttgaaca 3229
Db 5545 acggctgtgtcgcttcccaactccagccacagcgtgccttatgttctgaacttlttcgga 5604
OY 3240 agccaagaagcctgtgtccaaagaagctgtgcacacatagcgaacctacaaacagcctctcta 3299
Db 5605 ggtcgtgaagcgtgcagtggaagcagagccgttccctgtctcatatccctccagcctctgtc 5664
OY 3300 tgcacaagaagcccaaglatcacactgtgtctcggccggctgtgcgtgagagatgtgtgcgtgtgc 3359
Db 5665 tgcacagcagctgtgtgtcttccacactgtgtccatgtggtcgtgtgcgtgcgaatgtctccactgc 5724
OY 3360 ctaccgcactacgtatgtcctctcagaagctgtgtgtgcacaacgtgtgtgtgagatcgttagacta 3419
Db 5725 ccacctgtgtgttccctctgtgtgcgtgtgtgtgcgaatgtgtcgtgtgtgtgtgtgtgtgtgtgt 5784
OY 3420 cgtatccagggccacaagaagtgaaatgtgtgagatgtcttctgttaccggatggaagaatgt 3479
Db 5785 ggggtgcgcgcgaagaacactctcaagaacgtgtggaatgtcgtatcgtctcgtgtgtgcagaatgt 5844
OY 3480 gaactgcacactgcagaagcagactatgtgtgagatgtg---cttccctgtgcagtgtgaacct 3536
Db 5845 ggtcctgcagatgtccgaaatgtgtctgtgtgtgtgaaggtgatacgaacgcgtgtgaatgtgagagct 5904
OY 3537 gctgcaggtctcactatgtccttccctcgtcgtcaacaactctcgtacagaaggtgtgtgctt 3596
Db 5905 gctgtatgt 5964
OY 3597 ttccaagaagctcaagcccgagagacagcatltttgaacacactgtgtacactgttccatccog 3656
Db 5965 tgcacaatgcaccccaagcgt 6024
OY 3657 tgcagacactgtttgtgcacaagaacagtggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 3716
Db 6025 taagaacactctgt 6084
OY 3717 ggcacattgagcacaacctcaactaattgtcaacgtctccttlttaagaatgac-----ctt 3768
Db 6085 aaactgtgagctgtgcatgtccttcaacagccacctcttaagtgtcacaagccagccaggggaa 6144
OY 3769 gtcaatgtgtactcttctgtgagactatgtctgtggaagcccaactgtcattacattacagccag 3828
Db 6145 gttgtcttcggcccaactcaagcctcaagcctcaatcatcatatgacgtcaggtcgtgacaacag 6204
OY 3829 gaccacgtccacaagaagacca-gtttgtgtgaatgtgaatgtgtgtgtgtgtgtgtgtgtgtgtgt 3887
Db 6205 ttctctgtgcccctgt 6264
OY 3888 catgtccgcaatgtgaatctcatatatttctgtgaacctlttgagagctctcc 3941
Db 6265 catgtccttcaatgtcatcatcatcatgtctgtgcagagccctcctctgtgacccccc 6318
RESULT 10
AAC76373 standard: cDNA: 1377 BP.
XX
AC AAC76373:
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORF1928 polynucleotide sequence SEQ ID NO: 3855.
XX

QY	2228	gcacacatgaatgactcctcaacgaaggtagcgaatgacactttactaccttcgatatc	2287
Db	1291	gtctgcgtgatgatacccaaccctcgggggccgcgtgtgacaaccttaactctcgatcct	1350
QY	2288	cgggggaagtcgaggaagtgtgacatctcccaaatgcaccactgaagacaagccaag	2347
Db	1351	cgggggaagtcgaggagctgtgtcaataactccagcttgcaccaagtgtagtaaccaag	1410
QY	2348	gcgtgagaagaagaatgtatctctacaaccgcttaactcttcagagagaacgttgaaagctgcc	2407
Db	1411	gtgtgaagcagaagaatgtctctacaacc---tgccttcaagagaacttgaaagctgcc	1467
QY	2408	agaacctgtgacccgtgtgtgtatccaacccccaagctgctgcacatgcttaacttcagcag	2467
Db	1468	ggagagagtcgaagcctggtgtgatacagatcccca-----	1500
QY	2468	actgtcagagcctgcccctgtaggaagaccagatacacacgctgtataacaacccggagcatgtgcgcg	2527
Db	1501	-----gcttcgctctgtagagacacagatgcccgcgtgtataaaccggggctgtcgccttg	1552
QY	2528	atctgtacaaccacatgtaggacagctgacctatgtccacacccggcttcaaggagacacgctgcg	2587
Db	1553	atcagatactcggccaccccgagagctgtaaatgacaacacccgcttcaatggagcgcgtgtg	1612
QY	2588	agctctcgtgcg	2599
Db	1613	agatgtcgtgcg	1624
RESULT 12			
AA	AF87121		
ID	AAE87121	standard; DNA; 1577	BP.
AC	AAE87121:		
XX	26-MAR-2002	(first entry)	
DT	NOV10 coding sequence.		
DE			
XX	NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;		
KW	cardiovascular; casein kinase II phosphorylation site; contraception;		
KW	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KW	epidermal growth factor; cell development; apoptosis; cell adhesion;		
KW	growth migration; cell structure; motility; cancer; immune disorder;		
KW	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;		
KW	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KW	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KW	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16; ds.		
OS	Unidentified.		
XX			
XX			
EH	Key	Location/Qualifiers	
FT	CDS	387..1391	
FT	/*tag= a		
FT	/product= "NOV10"		
XX			
XX	MO200136638-A2.		
PN			
XX			
PD	25-MAY-2001.		
XX			
PF	17-NOV-2000; 2000MO-US31543.		
XX			
PR	19-NOV-1999; 99US-166336P.		
PR	29-NOV-1999; 99US-167785P.		
PR	08-MAR-2000; 2000US-187844P.		
PR	16-NOV-2000; 2000US-0715417.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
PI	Shimkets RA, Lichenstein H, Vernet C, Fernandes E;		
XX			
XX	WPI: 2001-648134/74.		
DR	P-PSDB: AAB83366.		
DR			

XX Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing e.g. cancer, inflammation and
PT immune disorders -
XX
XX Claim 9, Page 35, 141pp; English.

CC The sequence encodes the NOV10 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC Cytostatic; contraceptive; anti-inflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonists are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of Serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV5-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX Sequence 1577 BP; 424 A; 410 C; 342 G; 401 T; 0 other;

Query Match	11.0%	Score 517.4	DB23	Length 1577	
Best Local Similarity	78.8%	Pred. No. 4,66136			
Matches 617	Conservative	0	Mismatches 156	Indels 0	Gaps
QY 1508	ccctcaacatcagagcccccggatlaaagactcgggatcagcagagggccctccatcgtccaggttc	1567			
Db 571	ccctcgtactctctgttlaacagctlaaagactcgggacaatacgttttaatgcccaggttc	630			
QY 1568	ttgcgtctacgvtggtgtgtgcagagactgtgtgttgacaacccaataaagtcagacaca	1627			
Db 631	ttcgatcacacatgvtgcgcgcgcgcacagctgctcttggaaacccggaattgactcca	690			
QY 1628	gtgcacagacccctccaaaggagagccgaattccactctctgtctctcagacactgtgtta	1687			
Db 691	atgctactccctccaaaggagagcacaatgactctcgcgtctctccagacagcgtgtata	750			
QY 1688	taaacatlgagaggaaggtctctgtccagttacatctccagacccaatgtgcctcaccca	1747			
Db 751	taaacataagagccaagatcatatccagtcagtatcatcactactaattggttgttata	810			
QY 1748	ttatagacaagttgtcgtctcccaaaaacttgtctatccacccccaaagaatgctctggaca	1807			
Db 811	tcattagacaattgtctatctcccaaaaatttgttatactcccaagaacaactctggaca	870			
QY 1808	gggtctcgaacaaactctactacagtgtagcagcaacacacagatatactacaattcagcaat	1867			
Db 871	gaattctcgcaaaactctacgaactttgacacaacaatgtgctacatcaatttagcaact	930			
QY 1868	tgatacagactcagctctgtctgacagtcacatcactccatccacacccacagtcactg	1927			
Db 931	taatacagactcaggttctgtctagtgatctcaacagatcccatccacacccacagtcactc	990			

OY	1928	tctcttgcttaagcgaacaaagccttgaaagcttgccccagaagcagcagcttcctgt	1987
Db	991	ttctctggcccacgcagcaagccttcacgtccctactcgttgacaacacagagacttcgt	1050
OY	1988	tcaatcaagacacaagaagcagagctggaagtcttacctgaagttccaagtgatccgagact	2047
Db	1051	tcaaccaagaacacaagaagcagagctggaagtatttgaagttccatgtgatacgagatg	1110
OY	2048	ccaaggtcttlaagcttcaagacccctccaggtctctccctggaagacccctgaagctcag	2107
Db	1111	ccaaggtctttagctgtgatacttccacaatccacatgcgcctcggaagacccctgaagttcag	1170
OY	2108	agctgagtgctgaggtgtgtgaaacttgccagtgacatccgtgtgagctctttctaacaagaa	2167
Db	1171	agcttgagtgctgaaagtgtgtgagcttgccagtgacatccgtgtgactctcttcgatatggc	1230
OY	2168	tgtgagagttcataccgcgggagccttggttgaggtgggtgtgtgctatgagcttgaact	2227
Db	1231	ccctgagaagatltgtcagcgggagcctctctgttgacctgggtgtgtgctcctagcgatgact	1290
OY	2228	gctactcatgatctctcaacccaggttggtgcgagtgtgacaccttctactacatccgatattc	2287
Db	1291	gtctgctgattgatctcccaaccctcgggggcgctgtgtgacaccccttactacttcgatgtct	1350
OY	2288	cgg 2290	
Db	1351	cgg 1353	
RESULT	13		
ID	AAA57363		
AC	AAA57363 standard; DNA; 1522 BP.		
XX	AAA57363;		
DT	03-OCT-2000 (first entry)		
DE	DNA encoding a human hyaluronan-binding protein, designated WF-HABP.		
XX			
KW	Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;		
KW	proliferative condition; metastasis; inflammation; ischemia;		
KW	host defence dysfunction; immune surveillance dysfunction; arthritis;		
KW	multiple sclerosis; autoimmunity; immune dysfunction; allergy; ss.		
XX			
OS	Homo sapiens.		
XX			
FT	Key	Location/Qualifiers	
FT	CDS	31..1404	
FT		/*tag= a	
FT		/product= "hyaluronan-binding protein"	
FT		/transl_except= (pos: 1102..1104, aa: Xaa)	
FT		/transl_except= (pos: 1255..1257, aa: Xaa)	
FT		/transl_except= (pos: 1297..1299, aa: Xaa)	
FT		/note= "Xaa is an unspecified amino acid"	
XX			
PN	WO200039166-A1.		
XX			
PD	06-JUL-2000.		
XX			
PF	20-DEC-1999;	99WO-US30462.	
XX			
PR	23-DEC-1998;	98US-0113871.	
XX			
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(AMNA-) AMERICAN NAT RED CROSS.		
XX			
PI	Hastings GA, Liau G, Tsifrina E;		
XX			
DR	WPI: 2000-452376/39.		
DR	P-PSDB: AAY93911.		
XX			
PT	New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP,		
PT	OE-HABP and BM-HABP, useful for treating proliferative conditions,		

Dh 782 cgtataagacactcttcgtccctgctcattgaagagcttctgtgacacacatgacgtgagtg 841
OY 3713 gccggagacattgagcaccaccctcactaatgcaagctctccttctaagac----- 3765
Dh 842 gcccaaaccttgagcctgcatgctcctcaagccaccctcctaagtgcacaaagccagccag 901
OY 3766 -cttgcaatgtgacacttcttgaggaactgctgggaagcaactgctcattacag 3824
Dh 902 ggaagtgtctcgcgcgcactcagcctcagcctcattcatcagtgagcagcgcctgaca 961
OY 3825 ccaagacacagctccaccacagaacca-ggtttgtgatgtaagaatcattctgcatgtgg 3883
Dh 962 acagttcctctggccctctgtgcccacaggacagttgtgttaagcgtatcattgtgtgg 1021
OY 3884 acatctcgcgcgcgcacatgagatcctcattatatttctggaacttggagagctcctcc 3941
Dh 1022 acatcagtcctcaatgacatcattcatcagctctgacagccctcctctgcaacccc 1079

RESULT 14

AAS00854

ID AAS00854 standard; cDNA: 718 BP.

AC AAS00854;

DT 04-JUL-2001 (first entry)

XX Human cDNA clone HLMIT84 encoding cancer related protein 28.

KW Human; cancer related protein; HLMIT84; food additive;
KW preservative; immunogen; antibody; bone cancer; adrenal cancer;
KW bone marrow cancer; breast cancer; gastrointestinal cancer;
KW liver cancer; lung cancer; urogenital cancer; immune disorder;
KW Addison's disease; allergy; autoimmune haemolytic anaemia;
KW autoimmune thyroiditis; diabetes mellitus; Crohn's disease;
KW multiple sclerosis; rheumatoid arthritis; ulcerative colitis;
KW acquired immunodeficiency syndrome; AIDS; cardiovascular disorder;
KW myocardial ischaemia; wound healing; neurological disorder;
KW Parkinson's disease; Alzheimer's disease; cerebral anoxia; epilepsy;
KW viral infection; bacterial infection; fungal infection;
KW parasitic infection; agonist; antagonist; ss.

XX Homo sapiens.

OS Location/Qualifiers

FH Key 1..321

FT CDS /*tag= a

FT /product= "Cancer related protein 28"

FT /partial

FT /note= "No start codon"

XX MO200118014-A1.

XX PD 15-MAR-2001.

XX 30-AUG-2000; 2000WO-US23794.

XX PE 03-SEP-1999; 99US-0152296.

XX PR 06-OCT-1999; 99US-0158003.

XX (HUMA-) HUMAN GENOME SCI INC.

XX PA

XX XX

XX Roschke V;

XX PI

XX WPI: 2001-235186/24.

XX DR P-PSDB; AAU00893.

XX XX

XX PT

XX PT

XX PT

XX XX

XX PS

Disclosure: Page 383; 427pp; English.

CC The sequence encodes a novel Human cancer related protein. The
CC polynucleotides and polypeptides are useful for preventing,
CC treating or ameliorating a medical condition in e.g. humans,
CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities. The polynucleotide are
CC useful for chromosome identification. The nucleic acids, protein,
CC antibodies, agonists and antagonists are useful in the diagnosis,
CC treatment and prevention of cancer (e.g. cancers of the adrenal gland,
CC bone, bone marrow, breast, gastrointestinal tract, liver, lung, or
CC urogenital), immune disorders (e.g. Addison's disease, allergies,
CC autoimmune haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis), acquired immunodeficiency syndrome (AIDS), cardiovascular
CC disorders such as myocardial ischaemias, wound healing, neurological
CC diseases (e.g. Parkinson's disease, Alzheimer's disease, cerebral anoxia
CC and epilepsy) and infectious diseases such as viral, bacterial, fungal
CC and parasitic infections. Numerous examples of each type of disorder are
CC given in the specification.

XX Sequence 718 BP; 166 A; 206 C; 165 G; 181 T; 0 other;

SO

Query Match 3.8%; Score 177.4; DB 22; Length 718;

Best Local Similarity 63.9%; Pred. No. 1.2e-39;

Matches 334; Conservative 0; Mismatches 181; Indels 8; Gaps 4;

OY 3937 cctccacagcacaacagcgtgcaccactgtgcctgggagacgatattctgtccgtc 3996

Dh 10 cccctgtcccccgtgaccttgaccacacatgctgtggagacgatctcttgcacatc 69

OY 3997 gtctgtacactgtgcatgtctgtcgtcgtcacttacttccgtccgacaaagcagca 4056

Dh 70 atctcgtgtactggggtgtgtctgtcctgtgtccttaccctcacttctggataaccggaga 129

OY 4057 accactggttccagcgttttga-tcagaagagacatgtatgcttgccttggcaag 4115

Dh 130 acaatcgcttcacagcatttggatgctggaagagacattatgttcagctcttggcaag 189

OY 4116 cagaagcccaagaatctcgaacacccctctgtatgagacctcagccgacacccacagag 4175

Dh 190 cagagccttgagatattctgaacctctgtatgagacacacctcagctcccccagaa 249

OY 4176 tctctcgtgtaccccttcacagaccctgggagacagatctggagcagcgaccctctg 4225

Dh 250 ccttcctcagaccccttcacagaccctgtgaagacggagcttgagggaatgaccccttg 309

OY 4236 ggggcactg--cgctcgtgacatgagaagcagaacacacacacacagtggtccacg 4292

Dh 310 aggaactgtgagggccttgagcgggagatgacacatcactcactgacacttggccat 369

OY 4293 gtgattccagcccccagctgtctcatgcatcagtgattttaaagaatgacaacactcata 4352

Dh 370 caactgtgaattctcagc--accagttgcctttagaagacgtaaatgctttagaact 426

OY 4353 agcagccatacctccacccctctgttaattgttgatgttcgacagggctaagagccat 4412

Dh 427 cagaagcatatccctcatctctcgtgcatctgggggtgtgttcttcgtgggtgagagatgt 486

OY 4413 gtgccttgatccttgaggagccctccacccctcctcagacat 4455

Dh 487 gtgtgctgcccacacagatcagcttctcctcctgacccctt 528

RESULT 15

AAS62772

ID AAS62772 standard; cDNA: 722 BP.

XX AAS62772;

XX 14-FEB-2002 (first entry)

XX PT

XX PT

XX XX

XX DE cDNA sequence #559 encoding novel human secreted protein.

XX Human secreted protein; hyperproliferative disorder; autoimmune disorder;
KM Immune deficiency disorder; blood disorder; inflammatory disorder;
KM Infectious disorder; gene therapy; antimicrobial; hepatotropic;
KM Immunosuppressive; antineumatic; ss.
XX
OS Homo sapiens.
PN MO200177291-A2.
XX
PD 18-OCT-2001.
XX
PF 29-MAR-2001; 2001MO-US10485.
XX
PR 06-APR-2000; 2000US-195604P.
XX
PA (GEMV) GENETICS INST INC.
XX
PI Wong GG, Clark HF, Fechtel K, Agostino MJ, Howes SH, Resnick RJ;
PI Gulukota K, Graham JR;
XX
DR WPI; 2002-010900/01.
XX
XX
PT New polynucleotides encoding secreted proteins useful for treating e.g.
PT asthma, HIV and Crohn's disease -
XX
PS Claim 1: Page 358; 391pp; English.
XX
CC The present invention relates to the isolation of novel cDNA sequences
CC which encode human secreted proteins. The cDNA sequences have been
CC derived from a variety of human tissues. The invention also provides
CC a method for producing proteins from these polynucleotide sequences.
CC The proteins are useful for identifying compounds that modulate their
CC activity and production, and the cell is also useful for identifying
CC compounds that modulate expression of the polynucleotide sequences
CC encoding the secreted proteins. The sequences of the invention are
CC useful for treating diseases such as hyperproliferative disorders
CC (e.g. cancer), immune deficiency disorders (e.g. severe combined
CC immunodeficiency (SCID)), autoimmune disorders (e.g. multiple
CC sclerosis), blood disorders (e.g. thrombocytopaenia), inflammatory
CC disorders (e.g. arthritis) and infectious disorders (e.g. hepatitis).
CC The polynucleotide sequences of the invention are also useful in gene
CC therapy. AAS62214-AAS62838 represent the cDNA sequences of the
CC invention that encode for novel human secreted proteins.
XX
SQ Sequence 722 BP; 167 A; 197 C; 164 G; 194 T; 0 other;

Query Match 2.3%; Score 109.4; DB 24; Length 722;
Best Local Similarity 60.9%; Pred. No. 2.4e-20;
Matches 231; Conservative 0; Mismatches 141; Indels 7; Gaps 3;

QY 4080 tcagaagaagacatgtatgtcttggcatttggcaagacagaccccaagaatatcgcaaac 4139
 || ||||| ||||| ||| || ||||| ||||| ||||| ||||| |||
Db 174 tcggaagaagacataatgttcagctcttggcaagcagcagcctgagaatctcgaac 233

QY 4140 ccctctatagacactagcgccgacccccaagatcctcctgtgaccccttaacagac 4199
 || ||||| ||||| ||| || ||||| ||||| ||||| ||||| |||
Db 234 cccttgtatgagacacactccagctccccaagacactcttactgaccccttaacgagac 293

QY 4200 ccctgagaacagagatctggaggaacagcaccctctgggggacatg---cagtcctgacat 4256
 || ||||| ||| || ||||| || ||||| || ||||| || ||||| |||
Db 294 tctgaaagaacgacgagcttgagggacaatgaccccttgaggaacactgtgaggcctgagcgg 353

QY 4257 gagaagcccaagcaaacccacagtcacggttccacggtgaltcccaagcccacgtgtctc 4316
 || || ||||| || || ||||| || || ||||| || || ||||| ||
Db 354 gagatgcacagcactcactcactgcacactgagccatacactgtgaattctcagc---acc 410

QY 4317 atggatcagttgttttaagaatagacaacactataagaagccacatcaccctctctg 4376
 || || || || || || || || || ||||| ||||| |||
Db 431 agttgcctttagaagacgttaagctcttaagcactagaagccatacctcatctctctg 470

QY 4377 gtaaatctggatgtctgcagaggctaaagagacatgttgcctgataccctgggggacct 4436

Db 471 gctgatctgggggtgttcttctgtgtgagagatgtgttg--ctgtgccaccacagatacag 529
 || ||||| || || || || ||||| ||||| ||| ||
QY 4437 ccactctctgagcctat 4455
 || ||||| || |||
Db 530 ctctcctctgaccctt 548
 || ||||| || |||

Search completed: June 17, 2002, 16:44:47
Job time: 9872 sec

Tue Jun 18 09:06:50 2002

us-09-842-930a-1.rng

Page 24

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 13:20:30 : Search time 88.46 Seconds
(without alignments)
13067.502 Million cell updates/sec

Title: US-09-842-930A-1
Perfect score: 4706
Sequence: 1 tcttaccagaagctactcac.....aagcaaaaaaaaaaaaaa 4706

Scoring table:
IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2-6/ptodata/2/lna/5A.COMB.seq:*
2: /cgn2-6/ptodata/2/lna/5B.COMB.seq:*
3: /cgn2-6/ptodata/2/lna/6A.COMB.seq:*
4: /cgn2-6/ptodata/2/lna/6B.COMB.seq:*
5: /cgn2-6/ptodata/2/lna/PCTUS.COMB.seq:*
6: /cgn2-6/ptodata/2/lna/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	70.8	1.5	1414	1 US-08-024-868-1	Sequence 1, Appli
2	70.8	1.5	1414	2 US-08-242-097-1	Sequence 1, Appli
3	70.8	1.5	1414	4 US-09-206-695-1	Sequence 1, Appli
4	60	1.3	1400	2 US-08-001-078A-2	Sequence 2, Appli
5	60	1.3	1400	2 US-08-463-218-2	Sequence 2, Appli
6	60	1.3	1400	5 PCT-US94-00253-2	Sequence 2, Appli
7	44.2	0.9	7218	1 US-08-232-463-14	Sequence 14, Appli
8	43.4	0.9	7218	1 US-08-232-463-14	Sequence 14, Appli
9	41.2	0.9	1519	1 US-08-225-477B-2	Sequence 1, Appli
10	41.2	0.9	1519	5 PCT-US95-04353-2	Sequence 2, Appli
11	40.8	0.9	5191	1 US-08-340-428B-1	Sequence 1, Appli
12	40.8	0.9	5191	5 PCT-US93-07306-1	Sequence 1, Appli
13	40.4	0.9	8224	6 5180808-1	Patent No. 5180808
14	38.6	0.8	2461	1 US-08-282-141-1	Sequence 1, Appli
15	38	0.8	2049	1 US-08-268-797-1	Sequence 1, Appli
16	38	0.8	2049	5 PCT-US95-08414-1	Sequence 1, Appli
17	38	0.8	2691	1 US-07-878-960-1	Sequence 1, Appli
18	38	0.8	2892	1 US-08-264-534-5	Sequence 5, Appli
19	38	0.8	2892	1 US-08-083-5590A-1	Sequence 1, Appli
20	38	0.8	2892	1 US-08-465-500-5	Sequence 5, Appli
21	38	0.8	2892	2 US-08-346-126-5	Sequence 5, Appli
22	38	0.8	2892	2 US-08-346-126-5	Sequence 5, Appli
23	38	0.8	2892	3 US-08-532-384-1	Sequence 5, Appli
24	38	0.8	2892	3 US-08-893-828-5	Sequence 1, Appli
25	38	0.8	3259	5 PCT-US95-03747-1	Sequence 1, Appli
26	37.4	0.8	3215	1 US-08-426-627-1	Sequence 1, Appli
27	37.4	0.8	1520	1 US-08-225-477B-1	Sequence 1, Appli

28	37	0.8	1520	5 PCT-US95-04353-1	Sequence 1, Appli
29	36.4	0.8	3250	1 US-07-906-395-1	Sequence 1, Appli
30	36.4	0.8	3250	1 US-08-192-632-1	Sequence 1, Appli
31	36.4	0.8	3250	1 US-08-710-676-1	Sequence 1, Appli
32	36.4	0.8	3250	3 US-09-099-902B-1	Sequence 1, Appli
33	36.4	0.8	3250	5 PCT-US93-06080-1	Sequence 1, Appli
34	36.2	0.8	289	4 US-09-007-005-17	Sequence 17, Appli
35	36.2	0.8	289	4 US-09-244-796-17	Sequence 17, Appli
36	36.2	0.8	18609	4 US-08-943-731-1	Sequence 1, Appli
37	36	0.8	975	6 5340934-10	Patent No. 5340934
38	34.6	0.7	2447	2 US-09-014-969-14	Sequence 14, Appli
39	34.4	0.7	977	6 5340934-9	Patent No. 5340934
40	34.4	0.7	3709	4 US-09-541-782-7	Sequence 7, Appli
41	34.4	0.7	6049	4 PCT-US95-11684-3	Sequence 3, Appli
42	33.8	0.7	4411529	4 US-09-103-880A-1	Sequence 1, Appli
43	33.6	0.7	804	2 US-08-611-880-4	Sequence 4, Appli
44	33.2	0.7	198	1 US-08-330-108-16	Sequence 16, Appli
45	33.2	0.7	198	5 PCT-US92-10087-16	Sequence 16, Appli

ALIGNMENTS

RESULT 1
US-08-024-868-1
Sequence 1, Application US/08024868
Patent No. 5386013
GENERAL INFORMATION:
APPLICANT: Lee, Tae Ho
APPLICANT: Wisniewski, Hans Georg
TITLE OF INVENTION: Vllcek, Jan
TITLE OF INVENTION: Cytokine-Induced Protein, TSG-6, DNA
NUMBER OF INVENTIONS: 2
CORRESPONDENCE ADDRESSES:
ADDRESS: Browdy and Neimark
STREET: 419 Seventh Street, NW
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/024,868
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/642,312
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Livanat, Shmuel
REGISTRATION NUMBER: 33,949
REFERENCE/DOCKET NUMBER: Vllcek-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-628-5197
TELEFAX: 212-737-3528
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1414 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
CELL TYPE: Fibroblast
CELL LINE: FS-4
FEATURE:
NAME/KEY: CDS

Query Match	1.3%	Score 60;	DB 5;	Length 1400;
Best Local Similarity	52.7%	Pred. No. 3.3e-07;		
Matches 157;	Conservative 0;	Mismatches 135;	Indels 6;	Gaps 1.

Query Match 0.98; Score 44.2; DB 1; Length 7218;

```
Best Local Similarity 11.1%; Pred. No. 0.032;
Matches 52; Conservative 202; Mismatches 215; Indels 0; Gaps 0;

QY 857 ggcgaggagtgaaatgtgacatgagatcacacagacatgcaacggcgcgtgca 916
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1413 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1354

QY 917 ccaagtccaactgctcttgatccagcagcgaagcctcgtgcaaatgtgcgcagat 976
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1353 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1294

QY 977 tcgagggaaatgtaacggctgtgcacagcatalgtcctgtgagaaacgaatgagat 1036
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1293 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1234

QY 1037 gtctacaagaagcgactgtgataaagaaccccgaggaacgggtgtgtgtgcagc 1096
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1233 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1174

QY 1097 caggctataccgagcagcgtcgtgtgcttgaaatcaaccgtgtgtgagagacatg 1156
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1173 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1114

QY 1157 gtgagcttgaaagaatgcaagtgacacacagacagggcccaacgagcgctgtgta 1216
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1113 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1054

QY 1217 gcttcgcaaatatcacgtgagaatgtaagtgctgtcgtctatcaatgtgtcccaaga 1276
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1053 TCGACCTGCACCGCAACCTCGCAATTATTCTGTGACCTTATGCGCAACGAAAGAAATA 994

QY 1277 acaatgagcgctgcatcattgtcctctgtcgaactacactgacgaaga 1325
    || | || || | || | || | || | || | || | || | || | || | || |
Db 993 GTTATTAGACCGCATCGATGGACATTTCACGTAACGTTTAATA 945

RESULT 8
US-08-232-463-14
: Sequence 14, Application US/08232463
: Patent No. 5670367
: GENERAL INFORMATION:
: APPLICANT: DORNER, F.
: APPLICANT: SCHEIFLINER, F.
: APPLICANT: FALKNER, F. G.
: TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
: NUMBER OF SEQUENCES: 52
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Foley & Lardner
: STREET: 1800 Diagonal Road, Suite 500
: CITY: Alexandria
: STATE: VA
: COUNTRY: USA
: ZIP: 22313-0299
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM PC compatible
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/232.463
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/07/935.313
: FILING DATE:
: APPLICATION NUMBER: EP 91 114 300.6
: FILING DATE: 26-AUG-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: BENT, Stephen A.
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 30472/114 IMMU
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (703)836-9300
```

```
TELEFAX: (703)683-4109
:
: TELEX: 899149
: INFORMATION FOR SEQ ID NO: 14:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 7218 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: IMMEDIATE SOURCE:
: CLONE: pTZgt-F15
US-08-232-463-14

Query Match 0.9%; Score 43.4; DB 1; Length 7218;
Best Local Similarity 3.6%; Pred. No. 0.054;
Matches 14; Conservative 212; Mismatches 163; Indels 0; Gaps 0;

QY 4259 gaagcagaagaacacacagctcacggtcacggtatcccgccagcgtctcat 4318
    || | || | : : : : : : : : : : : : : : : : : : : : : :
Db 1055 GGAGCTTCGATYVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1114

QY 4319 ggaatcagttttaaagaatgacacacatcaagccagcataccctctcgtgt 4378
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1115 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1174

QY 4379 taatcggatgttcgcagcagcctaagagccatgttcctgatacctgggagcctcc 4438
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1175 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1234

QY 4439 acctcctctgagcctatacgtgtctctcctcactccatagtgctgtcgtctgc 4498
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1235 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1294

QY 4499 cctctctgtaccccaaacctgtactctgtgtatctctcatatgacgtgaacccaa 4558
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1295 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1354

QY 4559 ggcgggctcacctctatgtctgtatctcagtaaccagaagtaacctgcacacatgt 4618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1355 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1414

QY 4619 gtgcctaataatgtcttggaacaaat 4647
    : : : : : : : : : : | || || |
Db 1415 YVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVVV 1443

RESULT 9
US-08-225-477B-2
: Sequence 2, Application US/08225477B
: Patent No. 5635370
: GENERAL INFORMATION:
: APPLICANT: Susan Hockfield
: APPLICANT: Diane M. Jaworski
: TITLE OF INVENTION: BEHAV. A Brain Hya-
: TITLE OF INVENTION: Juruman-Binding Protein
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: St. Onge Steward Johnston & Reens
: STREET: 986 Bedford Street
: CITY: Stamford
: STATE: CT
: COUNTRY: United States
: ZIP: 06905
: COMPUTER READABLE FORM:
: MEDIUM TYPE: 3.5" 1.44 Mb diskette
: COMPUTER: IBM PC
: OPERATING SYSTEM: MS DOS
: SOFTWARE: Word Processor
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/225.477B
: FILING DATE: April 8, 1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Mary M. Krinsky
```



```
TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 202-628-5197
: TELEFAX: 202-737-3528
: INFORMATION FOR SEQ ID NO: 1:
:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 5191 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
:   MOLECULE TYPE: DNA
:
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 77..3847
:   US-08-340-428B-1
```

```
Query Match          0.9%: Score 40.8; DB 1; Length 5191:
Best Local Similarity 49.5%: Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
```

```
QY 3160 gacctctactccagagacagcagctagagatltccatcctcctcccactggccag 3219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 GACCTGTGTAACCCGTGGAGAGTACGGGCGTCTGTTCCATTATCGGGCGGCCGGACCGC 583

QY 3320 tacaactgacatttgacaagaagccaagaagcctgtgccaagaagcctgcagacataagc 3279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 TATCGGTGACCTTCGCGAGAGCCGAGAGCGCTTGACCTGACCTCCGCTACCATTTGCG 643

QY 3280 acctacaacagcctctcctatgcccagaagccaagatatacactgtgctgcgcgcctg 3339
    - - - - - ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 GCTCCAGACGACCTGCGACGCGCTGCTCGAAGATGCTTTGACAACTGCGATGCGGGCTGG 703

QY 3340 ctgagagatgagcgaggtgctcctaccgactac 3371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 CTCTCAGACCGCAGCGTCCGCTACCCGATCAC 735
```

```
RESULT 12
PCT-US93-07306-1
: Sequence 1, Application PC/TUS9307306
: GENERAL INFORMATION:
:   APPLICANT: MARGOLIS, Richard U.
:   APPLICANT: RAUCH, Uwe
:   APPLICANT: MARGOLIS, Renee K.
:   TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
:   NUMBER OF SEQUENCES: 48
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Broadway and Nelmark
:     STREET: 419 Seventh Street, N.W.
:     CITY: Washington
:     STATE: D.C.
:     COUNTRY: U.S.A.
:     ZIP: 20004
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patent Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: PCT/US93/07306
:     FILING DATE: 03-AUG-1993
:     PRIOR APPLICATION DATA:
:       APPLICATION NUMBER: US 07/922,911
:       FILING DATE: 03-AUG-1992
:     ATTORNEY/AGENT INFORMATION:
:       NAME: Townsend, Guy K.
:       REGISTRATION NUMBER: 34,033
:     REFERENCE/DOCKET NUMBER: Margolis-1A PCT
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 202-628-5197
:       TELEFAX: 202-737-3528
:     INFORMATION FOR SEQ ID NO: 1:
```

```
SEQUENCE CHARACTERISTICS:
:
: LENGTH: 5191 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA
:
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 77..3847
:   PCT-US93-07306-1
```

```
Query Match          0.9%: Score 40.8; DB 5; Length 5191:
Best Local Similarity 49.5%: Pred. No. 0.25;
Matches 105; Conservative 0; Mismatches 107; Indels 0; Gaps 0;
```

```
QY 3160 gacctctactccagagacagcagctagagatltccatcctcctcccactggccag 3219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 524 GACCTGTGTAACCCGTGGAGAGTACGGGCGTCTGTTCCATTATCGGGCGGCCGGACCGC 583

QY 3320 tacaactgacatttgacaagaagccaagaagcctgtgccaagaagcctgcagacataagc 3279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 584 TATCGGTGACCTTCGCGAGAGCCGAGAGCGCTTGACCTGAGCTCCGCTACCATTTGCG 643

QY 3280 acctacaacagcctctcctatgcccagaagccaagatatacactgtgctgcgcgcctg 3339
    - - - - - ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 644 GCTCCAGACGACCTGCGACGCGCTGCTCGAAGATGCTTTGACAACTGCGATGCGGGCTGG 703

QY 3340 ctgagagatgagcgaggtgctcctaccgactac 3371
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 704 CTCTCAGACCGCAGCGTCCGCTACCCGATCAC 735
```

```
RESULT 13
5180808-1
: Patent No. 5180808
: APPLICANT: RUOSLAHTI, ERKKI I.
:   TITLE OF INVENTION: VERSICAN CORE PROTEIN, NUCLEIC ACID
:   SEQUENCES ENCODING THE SAME, NUCLEIC ACID PROBES, ANTI-VERSICAN
:   ANTIBODIES, AND METHODS OF DETECTING THE SAME
:   NUMBER OF SEQUENCES: 4
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/07/441,179
:     FILING DATE: 27-NOV-1989
:     SEQ ID NO: 1
:     LENGTH: 8224
:   5180808-1
```

```
Query Match          0.9%: Score 40.4; DB 6; Length 8224:
Best Local Similarity 48.7%: Pred. No. 0.43;
Matches 110; Conservative 0; Mismatches 116; Indels 0; Gaps 0;
```

```
QY 3190 gttatccatctacgctctcccaactggccagatcacaaactgtgacaagccaagaag 3249
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 gtgttccactacgaggggagcaacagcaggtacacacgaattttaggcctcagaag 776

QY 3250 gctctgtgccaagaagcctgcagacataagccacttacaacacagctctcctatgccaagaag 3309
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 777 gctgtttgagcgtgtgggagcagctatagcaactccagagcagcctcttgcctcctaagaa 836

QY 3310 gccaaagtatacctgtgctgcgcgcgcgtgctgagagatgaggcggtgtgcctaccgact 3369
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 837 gatgatttgacagctgtgacgcagcgcgtgctgtgatacagactgtcatagatatccatc 896

QY 3370 acgtatgctctccagaagatgtgtgtgcaaaagttgttgatcgtaag 3415
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 897 cgggtctcccaagatgagctgtatagagataagatgggaaagcgag 942
```

```
RESULT 14
US-08-282-141-1
: Sequence 1, Application US/08282141
```

Patent No. 5538861
GENERAL INFORMATION:
APPLICANT: Schneider, Claudio
APPLICANT: Varnum, Brian
APPLICANT: Avanzoli, Giancarlo
APPLICANT: Brancolini, Claudio
APPLICANT: Manfioletti, Guido
TITLE OF INVENTION: Stimulating Factor for the AXL Receptor
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1840 Denavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: United States
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM/PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/282,141
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2461 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
US-08-282-141-1

```

Query Match          0.8%. Score 38.6: DB 1: Length 2461:
Best Local Similarity 52.1%: Pred. No. 0.69:
Matches 86: Conservative 0: Mismatches 79: Indels 0: Gaps 0:

Oy 2818 gtagcgggatacacatgcacagctgtggagcttcgaacaacagacaggggctgtgcg 2877
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 579 GGGGGCCGGCTTGGCGACAAGATGTCAACGATGCAAGCAGGCGAGAGAACGGGGCTGCGTC 638
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 2878 aagtcgcgtaatgtgtcccaagaagcaaccacagctctcttgacgtgcgaagaagctac 2937
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 639 CAGATCTGCGCAACAACAAGCGGGGTAGCTTCCACTGTCTCTCGCACAGCGCGCTTCGAGCTC 698
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Oy 2938 aaggggatgtgctacagctgcatalagatalagacccctgtgcagac 2982
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 699 TCCCTGATGGCAGGACCTGCCACAGCATAGACGATGCGCGAC 743
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
US-08-268-797-1
: Sequence 1, Application us/08268797
: Patent No. 5599788
:
: GENERAL INFORMATION:
: APPLICANT: Purchio, Anthony F.
: APPLICANT: LeBaron, Richard
: TITLE OF INVENTION: Factor to Grow Tissue Ex Vivo
: NUMBER OF SEQUENCES: 1
: CORRESPONDENCE ADDRESS:
: ADDRESSSEE: Knobb, Martens, Olson and Bear
: STREET: 620 Newport Center Drive, Sixteenth Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: USA
: ZIP: 92660
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/268,797
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: TISSUE.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ. ID NO.: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2049 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-268-797-1

```

	Query Match	0.8%	Score 38	DB 1	Length 2049
	Best Local Similarity	62.8%	Pred. No. 0.92		
	Matches 59	Conservative	Mismatches 35	Indels 0	Gaps 0
QY	1922	tcacgtctcttcgagcctgcagcaagccctcgcagagccttcgcccagagcagcagagact	1981		
Db	413	TCACCATCTTCGCCCTTCAGACAGCGAGCCCTGGCCCTCCTCTGCAGCTGAAGTGTGGACT	472		
QY	1982	tcctgttcacatcaagacaacaagaagcaagctaa	2015		
Db	473	CCCTGTGTCAGCATGTCAACATTGAGCTGCTCAA	506		

Search completed: June 17, 2002, 16:36:14
Job time: 11744 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd

OM protein - protein search, using sw model

```
Run on:      June 17, 2002, 12:26:09 ; Search time 41.98 Seconds
              (without alignments)
              3786.246 Million cell updates/sec
```

Title: US-09-842-930A-2
 Perfect score: 7861
 Sequence: 1 SLPSLLTRLEQMDYSIRG.....WGHCSPMRSQQAATTVPVR 1431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

1	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1980.DAT*
2	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1981.DAT*
3	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1982.DAT*
4	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1983.DAT*
5	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1984.DAT*
6	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1985.DAT*
7	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1986.DAT*
8	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1987.DAT*
9	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1988.DAT*
10	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1989.DAT*
11	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1990.DAT*
12	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1991.DAT*
13	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1992.DAT*
14	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1993.DAT*
15	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1994.DAT*
16	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1995.DAT*
17	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1996.DAT*
18	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1997.DAT*
19	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1998.DAT*
20	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA1999.DAT*
21	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2000.DAT*
22	/SIDS1/gcgdata/hold-geneseq/geneseqp-emb1/AA2001.DAT*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2813	35.8	2157	21	AAV93910	A human hyaluronan
2	2710.5	34.5	669	22	AAHB83364	NOV8 protein sequence
3	2701.5	34.4	669	22	AAHB83362	NOV6 protein sequence
4	1514.5	19.3	353	21	AAV93913	A human hyaluronan
5	1487	18.9	315	22	AAHB83358	NOV2 protein sequence
6	1318.5	16.8	545	22	AAHB83365	NOV9 protein sequence
7	1264.5	16.1	381	22	AAHB83363	NOV7 protein sequence
8	1209	15.4	244	22	AAHB83359	NOV3 protein sequence
9	1109.5	14.1	330	21	AAAB42164	Human OREF ORF1928
10	1067	13.6	334	22	AAHB83366	NOV10 protein sequence
11	774	9.8	457	21	AAV93911	A human hyaluronan

12	593.5	7.5	2189	11	AA805222	Antigen GX5401FL. e
13	496.5	6.3	2912	22	ABG06402	Novel human diagno
14	484	6.2	2703	22	AB860266	Drosophila melanog
15	469	6.0	2321	19	AAW49698	Human Notch3 prote
16	455.5	5.8	387	21	AA808932	Human secreted pro
17	453.5	5.8	3680	22	AB870878	Drosophila melanog
18	450	5.7	2471	20	AAV06816	Human Notch2 (hunn
19	441	5.6	4618	22	AAW39043	Human polypeptide
20	440	5.6	1872	19	AAW88510	Partial human Notc
21	438.5	5.6	1964	20	AAW85557	Mus musculus notch
22	431	5.5	3396	22	AB864261	Drosophila melanog
23	414.5	5.3	1404	22	AB861998	Drosophila melanog
24	413.5	5.3	1404	21	AAV59600	Drosophila Serrate
25	410	5.2	1193	17	AAW05885	Chick Serrate. Ga
26	410	5.2	1193	17	AAV59589	Chick Serrate prot
27	401.5	5.1	1036	18	AAW18331	Proliferation and
28	399.5	5.1	1010	20	AAW87896	Human JAGGED1 sol
29	399.5	5.1	1187	18	AAW18352	Proliferation and
30	399.5	5.1	1208	19	AAW00837	Human tagged prote
31	399.5	5.1	1218	17	AAW05883	Human Serrate-1 (H
32	399.5	5.1	1218	18	AAW18354	Proliferation and
33	399.5	5.1	1218	18	AAW44301	Human serrate 1.
34	399.5	5.1	1218	20	AAW87894	Human JAGGED1 prot
35	399.5	5.1	1218	21	AAV59597	Human Serrate prot
36	395.5	5.0	1404	14	AAW83830	Sequence of a serr
37	395	5.0	1118	22	AAW50209	Human fibrillin-li
38	378	4.8	1050	22	AAW66267	Human TANGO 272 SE
39	377.5	4.8	3088	21	AAW19794	Human laminin 2 ma
40	377.5	4.8	3089	21	AAW19792	Human laminin 2 ma
41	377.5	4.8	3110	16	AAW17130	Merosin major subu
42	377.5	4.8	3110	20	AAW15460	Human laminin alph
43	377.5	4.8	3110	21	AAW19793	Human laminin 2 al
44	377.5	4.8	3110	21	AAW19793	Human laminin 2 al
45	375.5	4.8	3150	22	ABG20414	Novel human diagno

ALIGNMENTS

RESULT	1
AAV93910	
ID	AAV93910 standard; Protein: 2157 AA.
XX	
AC	AAV93910;
XX	
DT	03-OCT-2000 (first entry)
XX	
DE	A human hyaluronan-binding protein, designated WF-HABP.
XX	
KW	hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;
KW	proliferative condition; metastasis; inflammation; ischemia;
KW	host defence dysfunction; immune surveillance dysfunction; arthritis;
KW	multiple sclerosis; autoimmunity; immune dysfunction; allergy.
XX	
OS	Homo sapiens.
XX	
Key	Location/Qualifiers
FT	344..353
FT	/note="cytochrome P450 cysteine haem-iron ligand binding domain"
FT	375..386
FT	/note="EGF-like type 1 domain"
FT	465..478
FT	/note="EGF-like type 2 domain"
FT	508..521
FT	/note="EGF-like type 2 domain"
FT	514..523
FT	/note="cytochrome P450 cysteine haem-iron ligand binding domain"
FT	551..564
FT	/note="EGF-like type 2 domain"
FT	943..954
FT	/note="EGF-like type 1 domain"
FT	Domain

FT	Domain	943..957	/note= "EGF-like type 2 domain"
FT	Domain	943..977	/note= "laminin-type EGF domain"
FT	Domain	987..998	/note= "EGF-like type 2 domain"
FT	Domain	987..998	/note= "EGF-like type 2 domain"
FT	Domain	987..998	/note= "EGF-like type 1 domain"
FT	Domain	1027..1040	/note= "EGF-like type 2 domain"
FT	Domain	1069..1082	/note= "EGF-like type 2 domain"
FT	Domain	1103..1113	/note= "prokaryotic membrane lipoprotein lipid attachment site domain"
FT	Domain	1111..1125	/note= "EGF-like type 2 domain"
FT	Domain	1405..1415	/note= "prokaryotic membrane lipoprotein lipid attachment site domain"
FT	Domain	1582..1616	/note= "laminin-type EGF domain"
FT	Domain	1582..1596	/note= "EGF-like type 2 domain"
FT	Domain	1582..1593	/note= "EGF-like type 1 domain"
FT	Domain	1626..1637	/note= "EGF-like type 2 domain"
FT	Domain	1626..1637	/note= "EGF-like type 1 domain"
FT	Domain	1663..1676	/note= "EGF-like type 2 domain"
FT	Domain	1747..1760	/note= "EGF-like type 2 domain"
FT	Domain	1791..1894	/note= "HA binding motif"
FT	Domain	1817..1862	/note= "link protein domain"
FT	Domain	1894..1908	/note= "EGF-like type 2 domain"
FT	Misc-difference	2058	/note= "unspecified amino acid encoded by Cnr"
FT	Misc-difference	2109	/note= "unspecified amino acid encoded by GAN"
FT	Misc-difference	2123	/note= "unspecified amino acid encoded by Tnr"
XX	WO200039166-A1.		
PD	06-JUL-2000.		
PF	20-DEC-1999;	99WO-US30462.	
PR	23-DEC-1998;	98US-0113871.	
PA	(HUMA-) HUMAN GENOME SCI INC.		
PA	(AMNA-) AMERICAN NAT RED CROSS.		
PI	Hastings GA, Llau G, Tsiftina E;		
DR	WP1: 2000-452376/39.		
DR	N-PSDB; AAA57362.		
PT	New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP, OE-HABP, and BW-HABP, useful for treating proliferative conditions, metastasis, inflammation, ischemia, arthritis and multiple sclerosis - Claim 11; Fig 1-H; 457pp; English.		

[illegible]

```
Dh 1600 afghbqacrtctvhyrdecdlgsgsfcfdegwlgtrpcvqlqlqvtcpcpapeavora 1659
Qy 929 NNTVCVCLNLESGGIRCTVVDVDFCKQNNNGSCAKYAKCSOKGQVSCSKKGKGCYCIE 988
Dh 1660 gnscecslygedqgrvtvadlclqdqngcscenhancsqvgtmvtctclpdyeqdwscre 1719
Qy 989 IDBCADGVNCGCEHATCRMTGPGKHKCEKSHYVGDVDC-EPEQLPLDCLQDNOCCH 1047
Dh 1720 rncptcdhrggscenhancletglttrccchagvygdqlcleseppvdrclqpppcc 1779
Qy 1048 PDASCADLYHQDTTVGVFHLRSPGLQYKLLFPDKAKECAKEAATIAITYNDLSVAOKAKYH 1107
Dh 1780 sdcmcltlhfgexravgfhlqatsgpyglhfseeaaceagavlasfpglsaaqqlgfh 1839
Qy 1108 LGSAGMLESRVAVPTTYASQCKGANYVGIYDVGSRANKSPMMVPCYRKMDVCTCKAG 1167
Dh 1840 LcImgwlaangslanpavfprvadcgngtrgvlsjgarhnlserwaycfrvadvacrcrnq 1899
Qy 1168 YVGDGFS-CSGNLLOVLMSPSLTNFLTEVLAFFSKSARGQAFLLKHLTDLISRGTFEPQ 1226
Dh 1900 fvgdglstcngkllldvlaacantstfygmllgyanacrqgldfidldelykltltv 1959
Qy 1227 NSGLPGKSLSGRDIIEHLTNVNSFYNDLVNGTFLRTMLGSQLLTFESQDLHQETRFV 1286
Dh 1960 negfvdmtlspnllehasnatllsan-asgqklpahsglslisldagpdmsswapva 2018
Qy 1287 DG-----SSLQWPIIANGLIHIIISPELRAP-----TAATAASGLGTGIFCAVLVLTG 1337
Dh 2019 pgtvsvsrllvwdlmalngllihalasplldppqavlaaxeappvaagavlaaqa1lg 2078
Qy 1338 AIALAASYFRLKORTTGPORF 1359
Dh 2079 lvaqalyllarqkpmgfgfsaf 2100

RESULT 2
AAB83364
ID AAB83364 standard; Protein; 669 AA.
XX
AC AAB83364;
XX
DT 26-MAR-2002 (first entry)
DE
DE NOV8 protein sequence.
XX
KW NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
KW cardiovascular; casein kinase II phosphorylation site; contraception;
KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
KW epidermal growth factor; cell development; apoptosis; cell adhesion;
KW growth migration; cell structure; motility; cancer; immune disorder;
KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;
KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.
XX
OS Unidentified.
XX
PN MO200136638-A2.
XX
PD 25-MAY-2001.
XX
PF 17-NOV-2000; 2000MO-US31543.
XX
PR 19-NOV-1999; 99US-166336P.
PR 29-NOV-1999; 99US-167785P.
PR 08-MAR-2000; 2000US-187844P.
PR 16-NOV-2000; 2000US-0715417.
XX
XX (CURA-) CURAGEN CORP.
PA
XX Shimkets RA, Lichenstein H, Vernet C, Fernandes E;
PI
DR MPI; 2001-648134/74.
```

```
DR N-PSDB; AAF87119.
XX
PT Novel human polypeptides and the nucleic acids that encode them useful
PT for preventing, diagnosing and treating e.g. cancer, inflammation and
PT immune disorders -
XX
PS Claim 1; Page 29-30; 141pp; English.
XX
CC This sequence is the NOV8 protein. The invention relates to
CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
CC cytosolic; contraceptive; antiinflammatory; immunomodulatory; and
CC cardiovascular activities. The sequences may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate NOVX
CC expression. They may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of protein by expressing inactive proteins or to
CC supplement the patients own production of protein. They are used to
CC produce NOVX proteins, by inserting the nucleic acid into a cell and
CC culturing it to express the protein. The DNA may be used as DNA probes in
CC assays to detect and quantitate the presence of similar DNAs in samples,
CC and which patients may need restorative therapy. The NOVX protein may
CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contraceptive). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
XX
SQ Sequence 669 AA;

Query Match 34.5%; Score 2710.5; DB 22; Length 669;
Best Local Similarity 78.8%; Pred. No. 1,5e-175;
Matches 471; Conservative 60; Mismatches 66; Indels 1; Gaps 1;
```

QY	929	NMTVCNLVVEGGGIGCTCTVDFEFCNNNGGCAKATYAKGCSQKTMQVSCSGCKYKGGYGCIE	988
Db	487	mtlceenllyegdgictctvdfekdnggacakarcesqgktkscscqykkgghnscete	546
QY	989	IDPCADGVNGGGEIHAATCRMTGPGKHRCCEKSHYVGDVCEPEEQLPLDRCLDQNGQCHP	1048
Db	547	idpcadglinggclghetackmttgyphknceckshyvgdglncepeqlidrcldqngqcha	606
QY	1049	DASCDALYRQDFVGVGFHILRSLPGLGVKLTFTDKKKEACAKANATITATYNOLSYAQKATY	1106
Db	607	dakcvalhtqdeevyvfhlrpslgykltfdkareacanaatmatynglsyaqkty	664
RESULT 3			
AAB83362	AAB83362 standard Protein: 669 AA.		
XX	AA	AAB83362;	
AC	XX		
DT	XX	26-MAR-2002 (first entry)	
DE	XX	NOV6 protein sequence.	
KW	XX	NOV: Cytostatic; contraceptive; antiinflammatory; immunomodulatory;	
KM	XX	cardiovascular; casein kinase II phosphorylation site; contraction;	
KM	XX	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;	
KM	XX	epidermal growth factor; cell development; apoptosis; cell adhesion;	
KM	XX	growth migration; cell structure; motility; cancer; immune disorder;	
KM	XX	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;	
KM	XX	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;	
KW	XX	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;	
KM	XX	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.	
OS	XX	Unidentified.	
XX	XX		
PN	XX	WO200136638-A2.	
PD	XX	25-MAY-2001.	
PE	XX	17-NOV-2000; 2000WO-US31543.	
PR	XX	19-NOV-1999; 990US-166336P.	
PR	XX	29-NOV-1999; 990US-167785P.	
PR	XX	08-MAR-2000; 2000US-187844P.	
PR	XX	16-NOV-2000; 2000US-0715417.	
PA	XX	(CUBA-) CUBAGEN GORP.	
PI	XX	Shimkets RA, Lichenstein H, Vernet C, Fernandes E;	
DR	XX	WPI: 2001-648134/74.	
DR	XX	N-PSDB: AAF87117.	
PT	XX	Novel human polypeptides and the nucleic acids that encode them useful	
PT	XX	for preventing, diagnosing and treating e.g. cancer, inflammation and	
PT	XX	immune disorders.	
PS	XX	Claim 1, Page 24-25; 141pp; English.	
CC	XX	This sequence is the NOV6 protein. The invention relates to	
CC	XX	the NOV1-NOV16 proteins, and their coding sequences. The proteins have	
CC	XX	cytostatic; contraceptive; antiinflammatory; immunomodulatory; and	
CC	XX	cardiovascular activities. The sequences may be used in the prevention,	
CC	XX	diagnosis and treatment of diseases associated with inappropriate NOVX	
CC	XX	expression. They may be used to treat disorders associated with decreased	
CC	XX	expression by rectifying mutations or deletions in a patient's genome	
CC	XX	that affect the activity of protein by expressing inactive proteins or	
CC	XX	suppress the patient's own production of protein. They are used to	
CC	XX	produce NOVX proteins, by inserting the nucleic acid into a cell and	
CC	XX	culturing it to express the protein. The DNA may be used as DNA probes in	
CC	XX	assays to detect and quantitate the presence of similar DNAs in samples,	
CC	XX	and which patients may need restorative therapy. The NOVX protein may	

CC also be used as antigens in the production of antibodies (Abs) against
CC NOVX and in assays to identify modulators of NOVX expression and
CC activity. The anti-NOVX Abs and antagonist are used to down regulate
CC expression and activity. The anti-NOVX Abs are used for detecting the
CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
CC characteristic of serine/threonine kinases, and are used to treat
CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
CC proliferation and contractions). NOV2-3, NOV6 and NOV8 are homologous to
CC the epidermal growth factor (EGF)-like super family and are involved in,
CC e.g., regulation of cell development, apoptosis, cell adhesion, growth
CC migration, cell structure and motility and protein management, and are
CC used to treat cancers, inflammatory disorders, immune disorders and
CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
CC fibrillin proteins and are used to treat cardiovascular disease e.g.
CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.

Query Match	34.48;	Score 2701.5;	DB 22;	Length 669;
Best Local Similarity	78.68;	Pred. No. 6.3e-175;		
Matches 470;	Conservative 60;	Mismatches 67;	Indels 1;	Gaps 1;

QY	RIKMDQGLMSQYLRFHYVAVGCCOOLLNDNLKVTTSATTLQGEPSISVSODIYAFINNEAK	568
Db	68 qvkwkdyglimpyvLrYhvaacqlllenkllsnatslgepervlsvsqvlyimnak	127
QY	VLSSDIISTNGVHVHDKLLSPKKLLTPKDALGRVLQNLTVAAANGYKRESKLIQDSG	628
Db	128 lssddistnglvhllskllspknllltlpkdnsgrrllgnltlatmgylkfssllqdsq	187
QY	LLSVITDSIHPRVTFVMPETKALEALRPBQDDPLFNDQNMKLSYLKSPHVIDSRALAS	668
Db	188 llsvltprlthprvclltpetqahlnaPraeqddllfngdmndkkeylkflvtldakvIav	247
QY	DLPRASAKWTLQSELSVRCGTSGDIELFLEQMCRFIRHGLFLFDGVAAYGIDCLLMP	748
Db	248 dlprstawkltlgelslvkcgagrdldgdlflngqtctlygrellfdlgvaygldclllldp	307
QY	TLGSRCDPTFTTIDPGEQSGCIFTPKCPKSKPKGVAKKCIYLPFRPRNVEGQMLCTV	808
Db	308 tlgrgcdrlftldasgegcscvmlprscprwskpkyvqkclyn-lpikrnllegcrrcsl	366
QY	VIGPRRCCHGVFMWDQACPEGRPTPCNNNGMCCDLYTPMGQCLCHTGFNGTACELCMHG	868
Db	367 vlgprckkygfgrcdqacp9p9p9p9nmgvclldqysatgcekcnglmgtaecmwp9	426
QY	RFGDQCPRSCSEHGQDDEGTLTSGSELCEFTGMAWASCDPTVAVFAVSPACGVNATCTE	928
Db	427 rfgrdclprcgshdgcdgdglfsgsgqclcecwmgp9scdtdqavlsvactprcsahatcke	486
QY	NNTCVCMNLNBYGDSITCTVYDFCKQNNNGCAKAAKCSOKGTQVSCSKKGTKGDYISCTE	968
Db	487 nntccnldyegddltcttvydfckqgdg9akavaresqkgltkvscsqkylgkdhnscte	546
QY	IDPCADGVNGGCHHATCRMTGPKKHKCECKSHVGVGVGVCPEPOLPLDNCLODNGGCHP	104
Db	547 idpraadgnggchhacckmtgprkhhkceckshvvgglncepeqplrdtclclqanqcha	606
QY	DASCADLYFDOTTVGVFHLRSPLOQVYKLTEDKAKEMACAEATLATYVNLQSLYAOKAKY	1106
Db	607 dakcvdlhfgqtlvgvfhllspglgqyltldtkareacanaeatmaeynqslsygqktyw	664
RESULT 4		
AAV93913		
ID	AAV93913 standard; Protein; 353 AA.	
XX		
AC	AAV93913;	
XT	03-OCT-2000 (first entry)	

Db	61	rsplqyqkltfkkareacaneaaetmatyqlslyxqakqnhlcsaayldeqrvayprlafas	120
Qy	1128	OKCGANVVGIVDYGSRANKSEMDVFCYRMKDVCNCTKAGYVGDDGFSGSGNLLQVLMSP	1187
Db	121	qncgsgvvgivdygpprpknksemwdfcygmkdvncetckvgyvgdgsysgnllqvlnmsfp	180
Qy	1188	SITNLTLEYLAFSKSKARQATFLKHLTDLTSLRGTLEVPONSGIPGKSKLSGRDIEHHLTN	1247
Db	181	slntfltevlaysnssargatflehltclslrgtlflvpqnsqigenetlsgriehhlan	240
Qy	1248	VNVSVYNDVNTGTLPTMTLSGSLTLTFESDOLH-OETRPVDGSLIOWDIIANGILHIT	1306
Qy	1307	SEPLRAPPTAAFAAHSGLTGTFICAVLVLTGAIAAAYSFRLKORTTGQFORF	1359
Db	241	vsmtfynndlvngtlctqltrfgskllltlrdqplhpletrcvedgtlewdicasngtlhvl	300
Db	301	srxklappaevrlnxhtgylgxlfxixllvtgavalaeyyfrlnrktlglfxhf	353
RESULT 5			
AAB83358	ID	AAB83358 standard; Protein; 315 AA.	
XX	AC	AAB83358;	
XX	DT	26-MAR-2002 (first entry)	
DE	XX	NOV2 protein sequence.	
XX	XX	NOV, Cytostatic; contraceptive; antinflammatory; immunomodulatory;	
KM	KM	cardiovascular; casein kinase II phosphorylation site; contraception;	
KM	KM	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;	
KM	KM	epidermal growth factor; cell development; apoptosis; cell adhesion;	
KM	KM	growth migration; cell structure; motility; cancer; immune disorder;	
KM	KM	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;	
KM	KM	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;	
KM	KM	therap; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;	
KM	KM	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.	
OS	XX	Unidentified.	
XX	PN	WO200136638-A2.	
XX	PD	25-MAY-2001.	
XX	PF	17-NOV-2000; 2000MO-US31543.	
XX	PR	19-NOV-1999; 990US-166336P.	
XX	PR	29-NOV-1999; 990US-167785P.	
XX	PR	08-MAR-2000; 2000US-187844P.	
XX	PR	16-NOV-2000; 2000US-0715417.	
XX	PA	(CURA-) CURAGEN CORP.	
PI	PI	Shimkets RA, Lichenstein H, Vernet C, Fernandes E;	
XX	DR	WPI: 2001-648134/74.	
XX	DR	N-PSDB: AAF87113.	
XX	PT	Novel human polypeptides and the nucleic acids that encode them useful	
XX	PT	for preventing, diagnosing and treating e.g. cancer, inflammation and	
XX	PS	immune disorders -	
XX	PS	Claim 1; Page 10-13; 141pp; English.	
CC	CC	This sequence is the NOV2 protein. The invention relates to	
CC	CC	the NOV1-NOV16 proteins, and their coding sequences. The proteins have	
CC	CC	cytostatic; contraceptive; antiinflammatory; immunomodulatory; and	
CC	CC	cardiovascular activities. The sequences may be used in the prevention,	
CC	CC	diagnosis and treatment of diseases associated with inappropriate NOVX	
CC	CC	expression. They may be used to treat disorders associated with decreased	
CC	CC	expression by rectifying mutations or deletions in a patient's genome	
CC	CC	that affect the activity of protein by expressing inactive proteins or to	

CC supplement the patients own production of protein. They are used to
 CC produce NOVX proteins, by inserting the nucleic acid into a cell and
 CC culturing it to express the protein. The DNA may be used as DNA probes in
 CC assays to detect and quantitate the presence of similar DNAs in samples,
 CC and which patients may need restorative therapy. The NOVX protein may
 CC also be used as antigens in the production of antibodies (Abs) against
 CC NOVX and in assays to identify modulators of NOVX expression and
 CC activity. The anti-NOVX Abs and antagonist are used to down regulate
 CC expression and activity. The anti-NOVX Abs are used for detecting the
 CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
 CC and/or treated vary depending on the NOVX protein. NOV1, NOV2, NOV5,
 CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
 CC characteristic of serine/threonine kinases, and are used to treat
 CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
 CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
 CC the epidermal growth factor (EGF)-like super family and are involved in,
 CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
 CC migration, cell structure and motility and protein management, and are
 CC used to treat cancers, inflammatory disorders, immune disorders and
 CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
 CC fibrillin proteins and are used to treat cardiovascular disease e.g.
 CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
 CC
 XX Sequence 315 AA;

Query Match 18.9%: Score 1487; DB 22; Length 315;

Best Local Similarity 78.3%: Pred. No. 6, 9e-93;

Matches 256; Conservative 24; Mismatches 31; Indels 16; Gaps 1;

QY 864 LCMHGRFGRDQPRSCSEHQDCDEGTGSGECLETGWTAAACDTPAVAVCTPACSVH 923
 DB 1 mcvgrfgrdclprcgshgqddgtltsqgcicetgwtgpcsdclavlsavctpcsa 60
 QY 924 ATCTENNTVCNNTNGSDGDTCTVNDPCKONNGCAKVAKCSKRGVSCSKKRGKGD 983
 DB 61 atktenntecndydgdtctvndpckvarkcsqkgtkvsccqkykkgd 120
 QY 984 YSCIEIDPCADGVNAGSGEHATCRMTGPRGKHKCECKSHYVGDVDEPEQLDPRDCLQDN 1043
 DB 121 hscieidpcadgvinggchehatckmtgprgkhhkceckshyvgdlinepeqlprdlclqdn 180
 QY 1044 GSCHPASCADLYFDQTVGVPHLRSPLGOYKLTFRKAEAKCAKEATATTWQLSYAK 1103
 DB 181 gscchadakadlfdqtlvghlrsplgoykltfrkareacaneaalmatlyqysyak 240
 QY 1104 AKNHLCSAGLSEGRVAPRTTVAOSKGANVGIIVGSRANKSEMDVVCYEMKQVNC 1163
 DB 241 aknhlcsagvlegrvaprataasqncsgsvvavdygprpnksmdvvcytmkg---- 296
 QY 1164 CKAGYVGDGFSGSLLOVLMSPSL 1190
 DB 297 -----saglfqgssrpsls 311

RESULT 6
 AAB83365
 ID AAB83365 standard; Protein; 545 AA.

AC AAB83365;

XX 26-MAR-2002 (first entry)

DE NOV9 protein sequence.

XX NOV9
 KW NOV, Cytostatic; contraceptive; antiinflammatory; immunomodulatory;
 KW cardiovascular; casein kinase II phosphorylation site; contraception;
 KW serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;
 KW epidermal growth factor; cell development; apoptosis; cell adhesion;
 KW growth migration; cell structure; motility; cancer; immune disorder;
 KW inflammatory disorder; cellular adhesion disorder; long-QT syndrome;
 KW cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;
 KW therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;

KW NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.
 XX Unidentified.
 OS WO200136638-A2.
 XX
 PN 25-MAY-2001.
 PD
 XX
 PF 17-NOV-2000; 2000WO-US31543.
 XX
 PR 19-NOV-1999; 99US-166336P.
 PR 29-NOV-1999; 99US-167785P.
 PR 08-MAR-2000; 2000US-187844P.
 PR 16-NOV-2000; 2000US-0715417.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 XX Shinketsu RA, Lichenstein H, Vernet C, Fernandes E;
 PI WPI: 2001-64834/74.
 DR N-PSDB; AAF87120.
 DR
 XX
 PT Novel human polypeptides and the nucleic acids that encode them useful
 PT for preventing, diagnosing and treating e.g. cancer, inflammation and
 PT immune disorders -
 PS
 XX Claim 1; Page 32-33; 141pp; English.

CC This sequence is the NOV9 protein. The invention relates to
 CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
 CC cystostatic; contraceptive; antiinflammatory; immunomodulatory; and
 CC cardiovascular activities. The sequences may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate NOVX
 CC expression. They may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of protein by expressing inactive proteins or to
 CC supplement the patients own production of protein. They are used to
 CC produce NOVX proteins, by inserting the nucleic acid into a cell and
 CC culturing it to express the protein. The DNA may be used as DNA probes in
 CC assays to detect and quantitate the presence of similar DNAs in samples,
 CC and which patients may need restorative therapy. The NOVX protein may
 CC also be used as antigens in the production of antibodies (Abs) against
 CC NOVX and in assays to identify modulators of NOVX expression and
 CC activity. The anti-NOVX Abs and antagonist are used to down regulate
 CC expression and activity. The anti-NOVX Abs are used for detecting the
 CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
 CC and/or treated vary depending on the NOVX protein. NOV1, NOV2, NOV5,
 CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
 CC characteristic of serine/threonine kinases, and are used to treat
 CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
 CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
 CC the epidermal growth factor (EGF)-like super family and are involved in,
 CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
 CC migration, cell structure and motility and protein management, and are
 CC used to treat cancers, inflammatory disorders, immune disorders and
 CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
 CC fibrillin proteins and are used to treat cardiovascular disease e.g.
 CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
 CC
 XX Sequence 545 AA;

Query Match 16.8%: Score 1318.5; DB 22; Length 545;

Best Local Similarity 75.7%: Pred. No. 4e-81;

Matches 240; Conservative 37; Mismatches 39; Indels 1; Gaps 1;

QY 509 RIRDMDOQGLMSOVLRXHVWGCOQLLDNLKVTTSATTLQGEVSVISVODTVFINNEAK 568
 DB 68 qvdkwdkylgmpvlylvhvachqlllenkllsnatslqgeplvsvsqtylmkak 127
 QY 569 VLSDDIITNGVHIVDKLISPKNLLITPDKADGRVLONTTVAANHGTYKFSKLIODSG 628
 DB 128 llsddistnglvhivdkllspknllitpdkdngsrllqnlctatnnykiksnllqdsq 187

OY	629	LLSVITDLSIHFFVYVFWPMDKALALEAPPQODLFFNODNKKDKLSYIKFHVIRDSKALAS	688
Db	188	llsvitldphhpvllfwpdqqlhalpaeqqdlflngdhkdklkeykllhvirdakylav	247
OY	689	DLPRASMKTLTGGSELTSVRCGTSGDIGELFNLNOMCRFIRHRCGLFDPGVAVYGIDCLLMP	748
Db	248	dlprstakwlltggseitsvrcgagrdgdlflngtqrtdvqrelltldgvaygidclltdp	307
OY	749	TLGRCODVFTFTFDIDPEGSCIFTPKCPKLSKPKGKKKCIYNPLPERRNVEGQNLCTV	808
Db	308	tlgsrcdtflttldasgecscvntpcspwskpkgykqkclyn-lpkkrnlqgcrercsl	366
OY	809	VIQTPRCCHGCFMPCDQ	825
Db	367	viqtpccckgylgrdcq	383
RESULT 7			
AAB83363			
ID	AAB83363	standard; Protein; 381 AA.	
AC	AAB83363;		
XX			
DT	26-MAR-2002	(first entry)	
XX			
DE	NOV7	protein sequence.	
XX			
KW	NOV; Cytostatic; contraceptive; antiinflammatory; immunomodulatory;		
KW	cardiovascular; casein kinase II phosphorylation site; contraception;		
KW	serine/threonine kinase; Peutz-Jeghers syndrome; cellular proliferation;		
KW	dermal growth factor; cell development; apoptosis; cell adhesion;		
KW	growth migration; cell structure; motility; cancer; immune disorder;		
KW	inflammatory disorder; cellular adhesion disorder; long-QT syndrome;		
KW	cardiovascular disease; hypertrophic cardiomyopathy; marfan syndrome;		
KW	therapy; NOV1; NOV2; NOV3; NOV4; NOV5; NOV6; NOV7; NOV8; NOV9; NOV10;		
KW	NOV11; NOV12; NOV13; NOV14; NOV15; NOV16.		
XX			
OS	unidentified.		
XX			
PN	MO200136638-A2.		
XX			
PD	25-MAY-2001.		
XX			
PF	17-NOV-2000; 2000MO-US31543.		
XX			
PR	19-NOV-1999; 99US-166336P.		
PR	29-NOV-1999; 99US-167785P.		
PR	08-MAR-2000; 2000US-187844P.		
PR	16-NOV-2000; 2000US-0715417.		
XX			
PA	(CURA-) CURAGEN CORP.		
XX			
II	Shimkels RA, Lichenstein H, Vernet C, Fernandes E;		
XX			
DR	WPI: 2001-648134/74.		
DR	N-PSDB: AAF87118.		
XX			
PT	Novel human polypeptides and the nucleic acids that encode them useful		
PT	for preventing, diagnosing and treating e.g. cancer, inflammation and		
PT	immune disorders -		
XX			
PS	Claim 1; Page 27-28; 141pp; English.		
XX			
CC	This sequence is the NOV7 protein. The invention relates to		
CC	the NOV1-NOV16 proteins, and their coding sequences. The proteins have		
CC	cytostatic; contraceptive; antiinflammatory; immunomodulatory; and		
CC	cardiovascular activities. The sequences may be used in the prevention,		
CC	diagnosis and treatment of diseases associated with inappropriate NOVX		
CC	expression. They may be used to treat disorders associated with decreased		
CC	expression by rectifying mutations or deletions in a patient's genome		
CC	that affect the activity of protein by expressing inactive proteins or		
CC	suppress the patients own production of protein. They are used to		

[illegible]

XX unidentified.
 OS
 PN WO200136638-A2.
 XX
 PD 25-MAY-2001.
 PF 17-NOV-2000; 2000MO-US31543.
 PR 19-NOV-1999; 99US-166336P.
 PR 29-NOV-1999; 99US-167785P.
 PR 08-MAR-2000; 2000US-187844P.
 PR 16-NOV-2000; 2000US-0715417.
 PA (CURA-) CURAGEN CORP.
 PI Shinkets RA, Lichenstein H, Vernet C, Fernandes E;
 DR WPI: 2001-648134/74.
 DR N-PSDB: AAF87114.
 XX
 PT Novel human polypeptides and the nucleic acids that encode them useful
 PT for preventing, diagnosing and treating e.g. cancer, inflammation and
 PT immune disorders -
 PT
 Claim 1: Page 14-17; 141pp; English.
 XX
 PS This sequence is the NOV3 protein. The invention relates to
 CC the NOV1-NOV16 proteins, and their coding sequences. The proteins have
 CC cytosolic; contractile; antiinflammatory; immunomodulatory; and
 CC cardiovascular activities. The sequences may be used in the prevention,
 CC diagnosis and treatment of diseases associated with inappropriate NOVX
 CC expression. They may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of protein by expressing inactive proteins or to
 CC supplement the patient's own production of protein. They are used to
 CC produce NOVX proteins, by inserting the nucleic acid into a cell and
 CC culturing it to express the protein. The DNA may be used as DNA probes in
 CC assays to detect and quantitate the presence of similar DNAs in samples,
 CC and which patients may need restorative therapy. The NOVX protein may
 CC also be used as antigens in the production of antibodies (Abs) against
 CC NOVX and in assays to identify modulators of NOVX expression and
 CC activity. The anti-NOVX Abs and antagonist are used to down regulate
 CC expression and activity. The anti-NOVX Abs are used for detecting the
 CC presence of NOVX in samples. Disorders that may be prevented, diagnosed
 CC and/or treated vary depending on the NOVX protein. NOV1, NOV3, NOV5,
 CC NOV7, NOV9-11 and NOV13-16 have casein kinase II phosphorylation sites
 CC characteristic of serine/threonine kinases, and are used to treat
 CC kinase-related disorders (e.g. Peutz-Jeghers syndrome, cellular
 CC proliferation and contraception). NOV2-3, NOV6 and NOV8 are homologous to
 CC the epidermal growth factor (EGF)-like super family and are involved in,
 CC e.g. regulation of cell development, apoptosis, cell adhesion, growth
 CC migration, cell structure and motility and protein management, and are
 CC used to treat cancers, inflammatory disorders, immune disorders and
 CC cellular adhesion disorders. NOV6-10 are homologous to EGF-like
 CC fibillin proteins and are used to treat cardiovascular disease e.g.
 CC hypertrophic cardiomyopathy, long-QT syndrome and marfan syndrome.
 CC
 XX Sequence 244 AA:
 SQ

Query Match 15.4%; Score 1209; DB 22; Length 244;
 Best Local Similarity 84.2%; Pred. No. 3.7e-74;
 Matches 202; Conservative 19; Mismatches 19; Indels 0; Gaps 0;

QY 864 LCMHGRFGDQPRSCSEHGQCEGTGSGCEITGCGMTAASCDPTAFAVACTPACSVH 923
 Db 1 mcvwprfipgdclpcqscsdgqcdgdgtltsqgclceqwgpsodtqavlsavctlpesah 60
 QY 924 ATCTENNTVCVNLNTEGDSITCTVDFGCKONNGCAKVAKCSOKGTQVSCSKCKGKGG 983
 Db 61 atckenntechaldyegddtictctvdftckdnggckavrcsqgkgtkvcscqgkykgdg 120

QY 984 VSCIEIDPCADSVNGSCHEHATCRMTGPCKHKCECKSHVGVGVCDEPRLDRLCLQDN 1043
 Db 121 hscleldpaddlmggchcalckmtlpgkhkceckshvvgdlncepeqlpdlcrlqdn 180
 QY 1044 GQCHPDASCADLYPDPTTGVGFHLRSPGQVYKLTFDKAEACAEATATATYNTQSTYAK 1103
 Db 181 gqchadackcvdlhfgdtlvvgvfhltspigqykltldkareacanaeatlnqstysqk 240

RESULT 9
 AAB42164
 ID AAB42164 standard. Protein: 330 AA.
 AC AAB42164;
 XX
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1928 polypeptide sequence SEQ ID NO:3856.
 XX
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vulnary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KW antiviral; antibacterial; antifungal; antineumatic; antihypoid;
 KW antineumatic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; nocturnal haemoglobinuria; burn; wound;
 KW thrombosis; contraceptive.
 OS Homo sapiens.
 XX
 XX MO200058473-A2.
 PD 05-OCT-2000.
 PF 31-MAR-2000; 2000MO-US08621.
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 PA (CURA-) CURAGEN CORP.
 PI Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB: AAC76373.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 PS Claim 11: Page 3007-3008; 5507pp; English.
 PS
 XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
 CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineumatic;
 CC antihypoid; and antineumatic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,

Sequence	457 AA;
XX	AA93911
XX	AA93911 standard; Protein; 457 AA.
XX	AA93911;
XX	03-OCT-2000 (first entry)
XX	
DE	A human hyaluronan-binding protein, designated WF-HABP.
KM	Hyaluronan-binding protein WF-HABP; OE-HABP; BM-HABP;
KW	proliferative condition; metastasis; inflammation; ischemia;
KM	host defence dysfunction; immune surveillance dysfunction; arthritis;
XX	multiple sclerosis; autoimmunity; immune dysfunction; allergy.
OS	Homo sapiens.
XX	
FH	Key
FT	Domain
FT	/note= "HA binding domain"
FT	Domain
FT	/note= "link domain"
FT	Domain
FT	/note= "EGF-like type 2 domain"
FT	Misc-difference 358
FT	/note= "unspecified amino acid encoded by CNT"
FT	Misc-difference 409
FT	/note= "unspecified amino acid encoded by GAN"
FT	Misc-difference 423
FT	/note= "unspecified amino acid encoded by TNT"
XX	
XX	W020003916-A1.
XX	
PD	06-JUL-2000.
XX	
XX	20-DEC-1999; 99MO-US30462.
XX	
XX	23-DEC-1998; 98US-0113871.
XX	
XX	(HUMA-) HUMAN GENOME SCI INC.
PA	(AMNA-) AMERICAN NAT RED CROSS.
XX	
P1	Hastings GA, Liao G, Tsifrina E;
XX	
XX	WPI: 2000-452376/39.
DR	N-PSDB; AAA57363.
XX	
P7	New hyaluronan-binding proteins, known as full-length WF-HABP, WF-HABP
P7	OE-HABP and BM-HABP, useful for treating proliferative conditions,
P7	metastasis, inflammation, ischemia, arthritis and multiple sclerosis -
XX	
PS	Claim 11; Fig 2A-B; 457pp; English.
XX	
CC	The present sequence represents a hyaluronan-binding protein. The
CC	specification describes four hyaluronan-binding protein, known as
CC	WF-HABP, WF-HABP, OE-HABP, and BM-HABP. The polypeptides are useful
CC	for treating diseases such as proliferative conditions, metastasis,
CC	inflammation, ischemia, host defence dysfunction, immune surveillance
CC	dysfunction, arthritis, multiple sclerosis, autoimmunity, immune
CC	dysfunction and allergy.
XX	
XX	

Query Match	9.8%	Score 774	DB 21	Length 457
Best Local Similarity	40.5%	Pred. No. 3e+4		
Matches 162	Conservative	56	Mismatches 170	Indels 12
				Gaps 5

```

0y 971 VSCGCKGKGYKQDGYSTCIEDIPCADGVNGGCGHEATRMFGPKNHCCEKSHSYVDGDVDC 1029
    |:::| |:::|:::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|
Db 2 vltc|pdyegdgwscrapmctd|nrgcgsehanc|stg|nttrtcehagyyvqdg|qcl 61
0y 1030 EPBQPLPDRCLQDNGGCGCHPPASCDALYFDQTTYGVGFHLASPLGQYKLTFTDRAKEKCAKEA 1089

```

Db 62 eesepvrcrlgpppchsamctdihfgekravghlqatspgylntfseeeaa; 121

QY 1090 ATITVYQLSYADKAKYHILCSAGWLESSRVAHYPTTYASQCGCANVGYIDVSRANKSEM 1149

Db 122 avlasfqljasaqqlfhgclmgwangstahpvrpvadcgngtrgylslyarhnlser 181

QY 1150 WDVCEYPMKDVNCTCAAGYAGDDFS-CSGNLIQVLNLFPSLINFLEVLAFSSARGOA 1208

Db 182 wdaycfrvqdvactcrngfrvgdylscngkllldvlaatanfstfgymllgyanaatrgld 241

QY 1209 FLKHLLDLSTGRLEFVFPONSGLPCKNSLSGRODIEHLLJTNVNSFYNDLVNFTGLFTMIGS 1266

Db 242 flldfidelcykrltfvupnegfvdmtltsgmlielhasnatllsan-aasgklllpahsgl 300

QY 1269 QLILTFESQDQLHGETRRVDG-----RSLQMDIILANGLIHILISEPLRAPR-----TAATA 1319

Db 301 slissdagpnnsswapyarpltvvsvstliwmdinafrgllhalaaspllappqavlaaxea 360

QY 1320 AHSGLTGTFCAVVLVYGALALAAVSFYRLAKRTTGFOFR 1359

Db 361 prvaagvavlaaagallgvlvaagalylaagrkpmgfrfsaf 400

RESULT	12
AAR05222	
ID	AAR05222 standard; protein; 2189 AA

AC AAR05222;

DT 02-AUG-1990 (first entry)

DE Antigen GX5401FL encoded by *Eimeria tenella* genomic DNA.

KW *Eimeria tenella*; antigen GX5401FL; antigen GX5401; avian coccidiosis.

OS *Eimeria tenella*

PN WO9000403-A.

PD 25-JAN-1990.

PF 05-JUL-1989; 89WO-US02918.

PR 05-JUL-1988; 88US-0215162.

PA (GENE-) GENEX CORP.

PI Anderson DM, McCandliss RJ, Strausberg SL, Strausberg RL;

DR WPI; 1990-051586/07.

XX

PT which binds with antibodies against avian coccidia, and

XX

It is encoded by an open reading frame contained within the sequence of clone 533 which was derived from an *E. tenella* genomic library screened with radioactively labelled cDNA encoding the GX5401 antigen. It is of about 250 Kd. It carries several repeated peptide sequences and is rich in cysteine residues. The open reading frame also encodes a potential signal sequence for protein secretion. Also new are an expression vector contg. cloned gene, and host cells transformed with the vector. The transformed cells are used in a vaccine to immunise birds against avian coccidiosis. By labelling the peptides, they can be used as a type-specific probe. May also be used in an assay to detect Ab against the coccidia. The Abs are used to identify transformed cells contg. the DNA.

50 Sequence 2189 AA;

```

Query Match          7.5%: Score 593.5: DB 11: Length 2189:
Best Local Similarity 24.1%: Pred. No. 4.5e-31:
Matches 293: Conservative 112: Mismatches 466: Indels 343: Gaps 71:

OY 132 IQKNBCNNNDITIVGEGCKCSQAP-----C-----PLETKPLRE--TRKCIYSIYM 178
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 464 megnacsdid-----ecseasteipencvntegstflaekpyelvdgcv-kidf- 515

OY 179 GKRSVFICQPOCVRTIITRACMLASLAHNAKPAPEYKMCALGTASWDVNGTGTCC 238
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 516 -----car-----gac--nslahckenpealtaic-----tcia 542

OY 239 GLGFMGTACTCTEGCKYIHCDQACSV-----HGRCGSGPLGSGDCDDVGMW- 287
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 gysgdta-----gq-----hcdiddeciaendclpadggicentvgstckcaaygqd 593

OY 288 GVKCDMEITTDNC-NGT--CHTSANCLDDPDGKASCACAGFRGNQVCTAINACETSMG 344
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 594 gusc-----fdidecangthnchasalc-lntggsfecacnagfsgnvgendvdecstdad 649

OY 345 GCSYKADCKRTTPGNRVCCVCKAGY--TGDCIVCLEINPCLLENHGGCDRAECTOTGPNQAV 403
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 650 dcgentlcnltv-gsflectcmagfeadaaklckdidecasgthlctclntla-gsft 707

OY 404 CNCLPKYTGDKVCSLINVCLTNNGCGSPFARCNTEEDQRI-CTCKPDTGDIYCRGS 462
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 708 cecnpvgfdgdkhcedvdcgqglhdcnvhaeeseadtlctckcglajysqeg----- 761

OY 463 IYGLPKNPSTSQYFPOLOEHAVRELACGPPTVFAPLSSFNHEPRIKDWQOGLMSQY 522
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 762 -hge-----ngqgdidecaqdal-----cgentvctlnprgsfe-----cavc 797

OY 523 LRYHVVCCOQLLDNLKWTTSATTL-----QGEPVSISSODTYFINNEA----- 567
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 798 egfavvagk-----lkqatsltcididecndasknltcatsadggscktagyscslp 851

OY 568 -----KVLSSDIISTNGVI--HVIDKLSPKNLLITPKDALGIVLOLTTVAANHIGT 618
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 852 gfgqgdghctdidecatqgvogeha-----lceenaggs--ynctlea--gylt 894

OY 619 KFSKLIQDSGLSVITDSIHFPVTVFWPTDK-----ALEALPPEQODELFNODMND 669
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 895 -----gqdgavgcididecaastavlpandatcvtlegsyftcecp-----gyrlhengct 944

OY 670 KIKSLTKFHVIRDSKALASDLPRASAMKTLQSGELSVRCGTGSD--IGELFLINQMCRTIH 728
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 945 k-----ldfctsekycnanasckenda-----gleaictchsgyegnge--geegcknid 991

OY 729 RGLLFDCVAVGIDCLMNPITLQGRCDIFFTYFDIPGEGSCIFTPKCPILKSPKGVKKKC 788
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 992 -----ec-----svgepckdf--gegvcdvdspsgfsiscatgfkik-- 1025

OY 789 IYNPLPFRNVEGCONLCTVVIQTPRCHGVFMPDQACPGGPDTPCNNRSCRDLYTPM 848
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1026 -----trst--cgdi-----decldgk-mnlc--apvg-----gltctntvsf 1058

OY 849 GOCCLHTGF--NGTACEILCMHGRFPCDQPNRSCSEHGOCDEBIT-----GSSECLCETGM 901
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1059 -tcscaagftgdglfctce-----didecataah-tcdpnatcvtntvsfegckegf 1107

OY 902 TAASCDTPFAVACTP--ACSVHATCTENN-----TCVCNLNVE--GGITCTVWDPFK 952
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1108 sgdg--htcddidecadpnlkcdtlhkgicqngtsgytcgrypslaaagftcdhndeca 1166

OY 953 QNNGCANVAKCSQKGTQVSCSKKGYKGDGYSCTEIDPCADGVNGGCHAEHATCRMTGPG 1012
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1167 agtarcgtsfctvdqsgyckceckngyrgsgedcvddecceadvh-tsehatctnt-eg 1224

OY 1013 KHKCECKSHYVGDVDCPEBPLDLRCLQDNGCQHDPDASCADLYRQDPTTVGVFHLRSPUG 1072
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1225 shctcnegygqdgkckektyvp--c--dnspcgnamc-----eatadynctckag 1273

```

```

OY 1073 QYKLTFDKAEACAK-EAATIATYNOLSYAOKAKYH-----LCSAGMLESGHAAVPTTYA 1126
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1274 -----yemkdgaacvdidecsgtlnhcdphadcsntdgsftctcgsygvglcdev---- 1325

OY 1127 SOKGCANVVGIVDYSRANKSEMDVFCYRMKDVN-----CTCKAGYVGDGFSG 1175
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1326 -decaghnag-----c-----dinavctnvgstfctceksqfegdghec 1363

OY 1176 SCNLLQVLSFSLNFLTLEVLAFSKSSARG-----QAFLEKHLIDLSTRGLFVPQNSG 1229
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1364 tekvljppqihcdswlactceetkqslrckevalpjkxevklcpdadlsacqelgews 1423

OY 1230 LPG-NKSLSGRDIE 1242
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1424 cpyvdnlnshrtae 1437

RESULT 13
ABG06402
ID ABG06402 standard; Protein; 2912 AA.
XX
AC ABG06402;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6393.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
OS Homo sapiens.
PN WO200175067-A2.
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB: AAS70589.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 36761; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (II) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human

```



```

OY 180 KRSVEFIGOPQCVRTIIRACWLASIAHNAKPAPBEVKMCAIGTASV-----WD 228
Db 327 -----atohdrvasfyc-----acpmgktgllchlddaecvsnphchedaictdn 369
OY 229 GYNGTGTCCGCGIGFNGTAC-----ETCTEKGVIHCDQACSCYH-GRC-SQGPLDGDSD 281
Db 370 pvngratctcpqgftggaodqdvdecsly-----anpchehlyrcvntqgsf--lcp 418
OY 282 CDVGWRGVKCDMEITTDNC-NGCTHTSANCILDPDGKASCKCAAFRCNGTCTA-INAC 339
Db 419 cgrgyltprcctdv--neclsgprctngatc-ldrtqgftclcmagf--tqlycevdidec 473
OY 340 ETS---NGGCSSTKADCKRTTPGNRVCCVCKAGYTGDIYC-LEINPCLNHHGGCDRNAECT 395
Db 474 gspcvnggv-----ckdrvngfs-ctcpsgftgs--tcqlvddecast--pcrnagkcv 523
OY 396 OTGPNQAVNCNLPKTYTGDKVCSLINVCLTNNGGCSFPAFCVTEODORI-----CTCKPD 451
Db 524 dq-pdyecrcaeqfeg-----tlcdnvddcsdpd-chhgtrcvdgiastscacp 573
OY 452 YTGDDGIVCRGSIYELPKNPSTSOYFFOLQEHAVRELAGPFTVAPL-SSSFNHEPRI 510
Db 574 yrglr--cesgv-decrsgpc-----rhgkcldlvdkylercpsgtgvcnevn 621
OY 511 KDMDOO---GLMSQ-VLRYHVAVCOQLL--LDNLKVTTSATTLQGEPEVSISSODIVF 562
Db 622 ddcasnpclfyvordglnrydcv-cqpgftgplcnveinecasspcqeg-----669
OY 563 INNEAKVLSDDIITNGVHVIDKLLSPKN---LLITPKDALGRVLONLTVAAHNGYT 618
Db 670 -----gscvdygengf-----rclcpqslpplclpshpcan-----epcsng-- 707
OY 619 KFSKLIDDS--GLTSYITDSIHPVTFWPTDKALEALPPEQODFLFNODNKKLKSYLK 676
Db 708 ----lcydapqgftfcvcepg-----wsgrpcsgsl-----733
OY 677 FHVIRDSKALASDLPRSASWKTLOGSELSVRCGTGSDIGELFLNEOMCRFIHRGLLDVG 736
Db 734 ----ard--acesqpraggtcssdmgfhtcpcpg-----763
OY 737 VAVGIDCLLMNPTL-----GGRCDFFTFDITPGECSGCIPTPKCPLKSPKGVK-KKCI 789
Db 764 -vgqrgcelispctpnpcchgrces-----apqg-----lpvcscpqgwgprc- 807
OY 790 YNPLPFRANVEGC-----ONLCTVVIOTPPRC-CH-GYFMPDCQACPGGPD-TPCNNR 838
Db 808 -----qgdvdecaqaprcpgrgicnlagstfctchgytgspscddindcnpclng 861
OY 839 GMCRD-----LYTPMG-----QCLCHTFNGTAC 862
Db 862 gscqgdvgsfscscplpgfagrcardvdeclsnpcpgjctcdhvasftctcpqyggtfnc 921
OY 863 ELICWHRRFPDOPRSCSEHGOCDEGITSGBELCTGWTAAASC-----DTPAVFA 914
Db 922 e-----qdlpdcspsscfnngtvcgv-nfsclicrpygvgancqheadpcisrclhng 975
OY 915 VCTPA-CSYHATCTENNTVCNLIYEGDGLTC-TVVDFCKO---NNGGCAKVAKCSQKGT 969
Db 976 vcsaahpgrfctclesft-----gpcqqlvdcwsirpcqngg-----rcvqtga 1020
OY 970 QVSCGCKKGKYG-----DGYSCIEIDP-----CADGVNG 998
Db 1021 y--clcpqwsgrlcdlrsljprcreaaqlyvrlcqlcgaqgcvdedsshycvcpqgrtg 1078
OY 999 G-----CHEHATCRMTGPGKHCKECKSHYVGDGVCDEPQQLPLDRCLQDNQ 1045
Db 1079 shcegevdpciaqpcqhggtcr-gymgymceclpnyngd--ncedd--vdecasq--p 1130
OY 1046 CHPDASCADLYFQDTTGVGVHLSRPLGOYKLTFDKAKKAEAKAEATATYNNQLSYAOKAK 1105
Db 1131 cqhngscldv-----arylcsppgtlgyvlcelinedd---1163
OY 1106 YHLCASAG-WLESG-----RVAVPTTASQKCGANVVGIVDYSRANKSE 1148

```

```

Db 1164 ---cgppldsqprclmngtclvvgfrcctcpqy-glireadi-----ne 1208
OY 1149 MMDVFCYR-----MKD-----VNCTCKAGYVG 1170
Db 1209 crsgachaahtrdcldqdpqggtfrcclhagfsg 1240

```

Search completed: June 17, 2002, 12:30:21
Job time: 252 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:30:24 ; Search time 16.67 Seconds
(without alignments)
3323.796 Million cell updates/sec

Title: US-09-842-930A-2
Perfect score: 7861
Sequence: 1 SLPSLTLRLQEMPDYSIFRG.....WGHCPDMRSQATVTVPR 1431

Scoring table: BLOSUM62
Gapop 10.0 , Capext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	510	2871	FBN1_MOUSE	O61554 mus musculus
2	508.5	2907	FBN2_MOUSE	O61555 mus musculus
3	497.5	2871	FBN1_HUMAN	P35555 homo sapien
4	496.5	2911	FBN2_HUMAN	P35556 homo sapien
5	486	2524	NOTC_XENLA	P21783 xenopus lae
6	484	2703	NOTC_DROME	P07207 drosophila
7	483.5	2871	FBN1_PIG	O61536 sus scrofa
8	482.5	2444	FBN1_HUMAN	P46531 homo sapien
9	481	2871	FBN1_BOVIN	P48133 bos taurus
10	473.5	2437	NOTC_BRARE	P46530 brachydantio
11	473	2318	NOTC3_MOUSE	O61982 mus musculus
12	446	2531	NTC1_MOUSE	O07008 rattus norv
13	445	2531	NTC1_RAT	O01705 mus musculus
14	436.5	5.6	NTC4_MOUSE	P31695 mus musculus
15	435	5.5	TENX_HUMAN	P22105 homo sapien
16	416.5	1408	SERR_DROME	P81168 drosophila
17	413.5	1376	CRBH_HUMAN	P82279 homo sapien
18	403	1064	FBP1_STRPU	P10078 strongyloce
19	377.5	4.8	LMA2_HUMAN	P24043 homo sapien
20	371	3051	YMX3_CAEEL	P34576 caenorhabdi
21	357.5	4.5	EGF_RAT	P07522 rattus norv
22	356	4.5	LMB2_MOUSE	O61292 mus musculus
23	355.5	4.5	LMA2_MOUSE	O60675 mus musculus
24	348.5	4.4	LMB1_MOUSE	P02469 mus musculus
25	348.5	4.4	CRB_DROME	P10040 drosophila
26	348	4.4	LMB2_RAT	P15800 rattus norv
27	348	4.4	LRP1_HUMAN	O07954 homo sapien
28	345	4.4	LMA5_HUMAN	O15230 homo sapien
29	344	4.4	LMB1_HUMAN	P07942 homo sapien
30	344	4.4	LMA1_MOUSE	P19137 mus musculus
31	343	4.4	LMA1_HUMAN	P25391 homo sapien
32	335	4.3	LMF2_CAEEL	O21313 caenorhabdi
33	334	4.2	LMB1_DROME	P11046 drosophila

34	333.5	4.2	1808	1	TENA_CHICK	P10039 gallus gall
35	332.5	4.2	1217	1	EGF_MOUSE	P01132 mus musculus
36	331	4.2	3712	1	LMA_DROME	O00174 drosophila
37	328	4.2	1207	1	EGF_HUMAN	P01133 homo sapien
38	328	4.2	1798	1	LMB2_HUMAN	P55268 homo sapien
39	326.5	4.2	1429	1	L112_CAEEL	P14585 caenorhabdi
40	324.5	4.1	2201	1	TENA_HUMAN	P24821 homo sapien
41	324.5	4.1	4543	1	LRPL_CHICK	P98157 gallus gall
42	315.5	4.0	1247	1	NIDO_HUMAN	P14543 homo sapien
43	314.5	4.0	1746	1	TENA_PIG	O29116 sus scrofa
44	313	4.0	1245	1	NIDO_MOUSE	P10493 mus musculus
45	301.5	3.8	3718	1	LMA5_MOUSE	O61001 mus musculus

ALIGNMENTS

```

RESULT 1
ID      FBN1_MOUSE      STANDARD:      PRT: 2871 AA.
AC      O61554; 060826;
DT      01-NOV-1997 (Rel. 35, Created)
DT      01-NOV-1997 (Rel. 35, Last sequence update)
DT      01-NOV-1997 (Rel. 35, Last annotation update)
DE      Fibrillin 1 precursor.
OS      FBN1 OR FBN-1.
OC      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=95130561; PubMed=7829516;
RA      Yin W., Germiller J., Sanguinetti C., Smiley E., Pangillan T.,
RA      Pereira L., Ramirez F., Bonaldi J.;
RT      *Primary structure and developmental expression of Fbn-1, the mouse
RT      fibrillin gene.*;
RL      J. Biol. Chem. 270:1798-1806(1995).
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=CD-1; TISSUE=Kidney;
RA      Ota K., Kumar A., Wada J., Liu Z., Kanwar Y.S.;
RL      Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC      -1- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
CC      THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE
CC      LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
CC      -1- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
CC      MICROFIBRILS (BY SIMILARITY).
CC      -1- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
CC      EGF-LIKE DOMAINS.
CC      -1- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; L29454; AAA56840.1; -
DR      EMBL; U22493; AAA64217.1; -
DR      HSSP; P35555; IAPJ.
DR      MGD; MGI:95469; Fbn1.
DR      InterPro; IPR000152; Asx_hydroxyl.
DR      InterPro; IPR000561; EGF-like.
DR      InterPro; IPR001881; EGF_CA.
DR      InterPro; IPR002212; TB.
DR      Pfam; PF00068; EGF_46.
DR      Pfam; PF00683; TB; 9.
DR      SMART; SM00179; EGF_CA; 42.
DR      SMART; SM00001; EGF_like; 4.

```


CC	entitles requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; L39790; AAA74908.1; -			
DR	EMBL; S69359; AAC60685.1; -			
DR	HSSP; P35555; IEMN			
DR	MGD; MGI:95490; Fbn2.			
DR	InterPro; IPR00152; Asx_hydroxyl.			
DR	InterPro; IPR000561; EGF-like.			
DR	InterPro; IPR001881; EGF_Ca.			
DR	InterPro; IPR001438; EGF_II.			
DR	InterPro; IPR002212; TB.			
DR	InterPro; IPR008822; Znf-C2H2.			
DR	Pfam; PF00008; EGF; 46.			
DR	Pfam; PF00683; TB; 9.			
DR	PRINTS; PR00010; EGFBL00D.			
DR	SMART; SM00179; EGF_CA; 43.			
DR	SMART; SM00001; EGF_Like; 3.			
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.			
DR	PROSITE; PS00022; EGF_1; 2.			
DR	PROSITE; PS01186; EGF_2; 36.			
DR	PROSITE; PS01187; EGF_CA; 43.			
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain; Repeat; Signal; Multigene family.			
KW	Repeat; Signal; Multigene family.			
FT	CHAIN	1	28	POTENTIAL.
FT	DOMAIN	29	2907	FIBRILLIN 2.
FT	EGF-LIKE 1	111	142	EGF-LIKE 1, NON-CALCIUM BINDING.
FT	EGF-LIKE 2	145	176	EGF-LIKE 2, NON-CALCIUM BINDING.
FT	DOMAIN	176	208	EGF-LIKE 3, NON-CALCIUM BINDING.
FT	DOMAIN	276	317	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	318	359	EGF-LIKE 5, CALCIUM-BINDING.
FT	REPEAT	360	426	TGFBR 1.
FT	DOMAIN	487	527	EGF-LIKE 6, NON-CALCIUM BINDING.
FT	DOMAIN	528	567	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	568	609	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	610	650	EGF-LIKE 9, CALCIUM-BINDING.
FT	DOMAIN	651	691	EGF-LIKE 10, CALCIUM-BINDING.
FT	DOMAIN	692	760	TGFBR 2.
FT	REPEAT	761	802	EGF-LIKE 11, CALCIUM-BINDING.
FT	DOMAIN	803	844	EGF-LIKE 12, CALCIUM-BINDING.
FT	DOMAIN	845	883	EGF-LIKE 13, CALCIUM-BINDING.
FT	DOMAIN	948	989	EGF-LIKE 14, CALCIUM-BINDING.
FT	REPEAT	990	1065	TGFBR 3.
FT	DOMAIN	1066	1107	EGF-LIKE 15, CALCIUM-BINDING.
FT	DOMAIN	1108	1150	EGF-LIKE 16, CALCIUM-BINDING.
FT	DOMAIN	1151	1192	EGF-LIKE 17, CALCIUM-BINDING.
FT	DOMAIN	1193	1234	EGF-LIKE 18, CALCIUM-BINDING.
FT	DOMAIN	1235	1275	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN	1276	1317	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN	1318	1359	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN	1360	1400	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN	1401	1441	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN	1442	1483	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN	1484	1524	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN	1525	1565	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT	1566	1642	TGFBR 4.
FT	DOMAIN	1643	1684	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1685	1726	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT	1727	1800	TGFBR 5.
FT	DOMAIN	1801	1842	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1843	1884	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1885	1926	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1927	1965	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1966	2008	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	2009	2048	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2049	2090	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT	2091	2163	TGFBR 6.
FT	DOMAIN	2164	2205	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2206	2245	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2246	2286	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2287	2330	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN	2331	2372	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT	2373	2441	TGFBR 7.

FT	DOMAIN	2442	2483	EGE-LIKE 41, CALCIUM-BINDING
FT	DOMAIN	2484	2524	EGE-LIKE 42, CALCIUM-BINDING
FT	DOMAIN	2525	2563	EGE-LIKE 43, CALCIUM-BINDING
FT	DOMAIN	2564	2606	EGE-LIKE 44, CALCIUM-BINDING
FT	DOMAIN	2607	2646	EGE-LIKE 45, CALCIUM-BINDING
FT	DOMAIN	2647	2687	EGE-LIKE 46, CALCIUM-BINDING
FT	DOMAIN	2688	2727	EGE-LIKE 47, CALCIUM-BINDING
FT	DISULFID	115	124	BY SIMILARITY.
FT	DISULFID	119	130	BY SIMILARITY.
FT	DISULFID	132	141	BY SIMILARITY.
FT	DISULFID	143	159	BY SIMILARITY.
FT	DISULFID	159	164	BY SIMILARITY.
FT	DISULFID	166	175	BY SIMILARITY.
FT	DISULFID	180	190	BY SIMILARITY.
FT	DISULFID	184	196	BY SIMILARITY.
FT	DISULFID	198	207	BY SIMILARITY.
FT	DISULFID	280	292	BY SIMILARITY.
FT	DISULFID	287	301	BY SIMILARITY.
FT	DISULFID	303	316	BY SIMILARITY.
FT	DISULFID	322	334	BY SIMILARITY.
FT	DISULFID	329	343	BY SIMILARITY.
FT	DISULFID	345	358	BY SIMILARITY.
FT	DISULFID	491	503	BY SIMILARITY.
FT	DISULFID	498	512	BY SIMILARITY.
FT	DISULFID	514	526	BY SIMILARITY.
FT	DISULFID	532	542	BY SIMILARITY.
FT	DISULFID	537	551	BY SIMILARITY.
FT	DISULFID	553	566	BY SIMILARITY.
FT	DISULFID	572	584	BY SIMILARITY.
FT	DISULFID	579	593	BY SIMILARITY.
FT	DISULFID	595	608	BY SIMILARITY.
FT	DISULFID	614	625	BY SIMILARITY.
FT	DISULFID	620	634	BY SIMILARITY.
FT	DISULFID	636	649	BY SIMILARITY.
FT	DISULFID	655	666	BY SIMILARITY.
FT	DISULFID	661	675	BY SIMILARITY.
FT	DISULFID	677	690	BY SIMILARITY.
FT	DISULFID	765	777	BY SIMILARITY.
FT	DISULFID	772	786	BY SIMILARITY.
FT	DISULFID	788	801	BY SIMILARITY.
FT	DISULFID	807	819	BY SIMILARITY.
FT	DISULFID	814	828	BY SIMILARITY.
FT	DISULFID	830	843	BY SIMILARITY.
FT	DISULFID	849	859	BY SIMILARITY.
FT	DISULFID	854	868	BY SIMILARITY.
FT	DISULFID	870	883	BY SIMILARITY.
FT	DISULFID	952	964	BY SIMILARITY.
FT	DISULFID	959	973	BY SIMILARITY.
FT	DISULFID	975	988	BY SIMILARITY.
FT	DISULFID	1070	1082	BY SIMILARITY.
FT	DISULFID	1077	1091	BY SIMILARITY.
FT	DISULFID	1093	1106	BY SIMILARITY.
FT	DISULFID	1112	1124	BY SIMILARITY.
FT	DISULFID	1119	1133	BY SIMILARITY.
FT	DISULFID	1135	1149	BY SIMILARITY.
FT	DISULFID	1155	1167	BY SIMILARITY.
FT	DISULFID	1162	1176	BY SIMILARITY.
FT	DISULFID	1178	1191	BY SIMILARITY.
FT	DISULFID	1197	1209	BY SIMILARITY.
FT	DISULFID	1204	1218	BY SIMILARITY.
FT	DISULFID	1220	1233	BY SIMILARITY.
FT	DISULFID	1239	1250	BY SIMILARITY.
FT	DISULFID	1246	1259	BY SIMILARITY.
FT	DISULFID	1261	1274	BY SIMILARITY.
FT	DISULFID	1280	1292	BY SIMILARITY.
FT	DISULFID	1287	1301	BY SIMILARITY.
FT	DISULFID	1303	1316	BY SIMILARITY.
FT	DISULFID	1322	1334	BY SIMILARITY.
FT	DISULFID	1329	1343	BY SIMILARITY.
FT	DISULFID	1345	1358	BY SIMILARITY.
FT	DISULFID	1364	1377	BY SIMILARITY.
FT	DISULFID	1371	1386	BY SIMILARITY.
FT	DISULFID	1388	1399	BY SIMILARITY.

Query Match
Best Local Similarity 23.5%, Score 508.5; DB 1; Length 2907;
Matches 296; Conservative 123; Mismatches 395; Indels 445; Gaps 87;

FT DISULFID 1405 1418 BY SIMILARITY.
FT DISULFID 1412 1427 BY SIMILARITY.
FT DISULFID 1429 1440 BY SIMILARITY.
FT DISULFID 1446 1458 BY SIMILARITY.
FT DISULFID 1453 1467 BY SIMILARITY.
FT DISULFID 1469 1482 BY SIMILARITY.
FT DISULFID 1488 1499 BY SIMILARITY.
FT DISULFID 1494 1508 BY SIMILARITY.
FT DISULFID 1510 1523 BY SIMILARITY.
FT DISULFID 1529 1540 BY SIMILARITY.
FT DISULFID 1535 1549 BY SIMILARITY.
FT DISULFID 1551 1564 BY SIMILARITY.
FT DISULFID 1647 1659 BY SIMILARITY.
FT DISULFID 1654 1668 BY SIMILARITY.
FT DISULFID 1670 1683 BY SIMILARITY.
FT DISULFID 1689 1701 BY SIMILARITY.

Query 190 QCVRIITRAQMIASLANAKPAPG---EVKMCALGTASVMDGVNGTGT---CQCGLGFGNG 244
Db KQIPVSSYRC-RQNMGYK-QDANGDCIDVDECTSNPCSNQCVNTPSSYYCKCHAGFOR 559
QY 245 T---ACETTEGKXGHHQDO-ACSCVHGRCSQGLGDGS---CDQDVGMRGVKCDMEITTT 297
Db TPTKQACIDIDE-----CIONGVLCRNGRCVN---SDGSPQICINAGF-----ELTT 603
QY 298 D--NC--NGTHTSANCIL---LDPDGKASCKCAAGF---RGNGTVCIAINACT---SN 343
Db DQKNCVDHDECTTTNMCLNGMCINEDGSEFKCYCKPGFLLAPRGYCTDVEDCQTPGICMN 663
QY 344 GGC---STRADCKRTTPG-----NRVCY----- 363
Db GHCINNBSGFRDC---PRGLAVGVDRVCYDTNMRSTCYGEIKKGVCRPREGAVTKSE 720
QY 364 ----- 363
Db CCCANPDYGFGEPOPCPAKNSAEFHGLCSSGIGTVNDRDINECALDPDICANGICENL 780
QY 364 -----CKAGYTD--GIYCLEINPCLENHGGCDRNAECTOTGPNQAVCMCLPKYT--G 412
Db RGTFRNCNSGSEYBPASGRNICIDIDECLNRLICD--NGLCRNT--PGSYSCCTCPREYVLP 838
QY 413 DQKYSILNVLCTN---NGGSPRAFCNVTEDDQRTCTCKP--DYTGGIYCKRSYTG-- 465
Db ETECEDVNECESNFCVAGACR-----NMLGSEFHCECSPSKSLSTGLICIDISLAKTIC 891
QY 466 -----ELPKNPSTQYFFOLOEHAHVRELACGPFVFAPISSSNHPRIDMDQO 516
Db WLNIONDNCEVNINGAT-----LKSECCATLGA-----WSP 924
QY 517 GLMSOVLRYHVGCCQQLLDNLKVTTSATTLQGEPVSTVSODIVFINNEAKVL-----S 571
Db -----CERCELD-----AACPRGFARIKGVICEDV---NECEVFPFGVOPN 961
QY 572 SDIISTNCVHIVIDKLSPKULLITPKALGRVLONLTTVAANHCYTFPSKLIODSGLIS 631
Db 962 GRVNSKSKSFH---CECPBELTL--DGTGRVCLD---IRMEHCFLEKWD---EDECIHP 1008
QY 632 V-----ITDSIHTPYTVFMPDT-----KALEALPREQODFLFNQNNKDKLKYLF- 677
Db 1009 VPGAFRMDACCAGVAMGTCCECPKPGTKYEFTLCRPGFA---NRGDILTRGPRFC 1064
QY 678 HIVRDSKALASLDRPSASWKTLOGSELSVRCGTG-----SDIGELFLNEQMCRFI 727
Db 1065 KDINECKAFPCPMCTYTGKCRNLI--GSFKRCRNNGFALDMEERNCTDIDECRISPLC--- 1119
QY 728 HRGLLPDVGVAYGIDC-----LLM-----NPTL--GGRC-DPFTTFDIDIGE 765
Db 1120 GSGICVNTPGSFCECEGEYSGFMMKNCMDIDECERNPLLCRGTCGVNTEGSP---Q 1175

QY 766 GCSGIEFTKCPKLSKPKGVKKCI--YNPLPRRRNVEGCON-LCTVVIQTTPRC-CH-GY-F 820
Db 1176 C-----DCPLGHLSLSPREDCVDINBESLSDNL--CENKGVMMIGTYQSCNPGYA 1226
QY 821 MPDQACP-----GGPDTPCNNRGMCRDLYTPMGQCLCHTGP---NCTAC---EL 864
Db 1227 TPDROGCTDIDECMIMNGGCDTQCTN---SEGSY---ECSCEGYALMPDGRSCADIDE 1279
QY 865 CMHGRFPGDQCRSRSEHQDCBEGITGSEGCICETGWTAAASCDTPFAV-FAYC--TPACS 921
Db 1280 C-----ENNDPIC-DGGQC--TNIPGEYRCLCYDGF-MASMDMTCTCIDVNECDLNPNIC 1329
QY 922 VVATCTENN-----TCVCNLNVE-GDGIT-CTVVDPCCKONNGCAVAKOSKGTQVSCS 975
Db 1330 MEGEC-ENTKSGSFICHQGLGYVKKGTCTCTTIDECCEICAHNCMDHNASLNPVGSFKSC 1388
QY 976 KKGKGDGYSCIEIDPCADGVNGGCHENATCRMTGPKHKKCKSKSHYVGDVDCPEBOLP 1035
Db 1389 REGWNGNGIKCIDLDECANGTI-QCSINAQCVNT--PGSYRCACSEGFQDGFTC---D 1442
QY 1036 LDRC-----LDNNGC-----HPDASCADL---YQDITV----- 1062
Db 1443 VDECAENTNLCENGQCLLVNPGAYRCECEMGTTPASDSRSCQDIDECSPQNTICVFQTCNNL 1502
QY 1063 -GVFHLRSPLOQYKLTFPKAKEACA---KEATTIATYNOLSYAOKAKYHLSAGMLESGR 1118
Db 1503 PGMFHCICDDG-YEL--DRTGNCCTDIDECADPINCYNGLCVNTPGRRE-CMC----- 1551
QY 1119 VAVPTTYASQKCGANVGVIVDYGSRANKSEMDVFCYRMKMDVNCCTOKAGYVGD--FSCS 1176
Db 1552 ---PDEF---QLNATGVGVD--NRVG-----NCYLKFGPRDGSLSGN 1587

RESULT 3
FBN1_HUMAN
AC P35555;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fibrillin 1 precursor.
GN FBN1 OR FBN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=93372860; PubMed=8364578;
RA Pereira L.V., D'Alessio M., Ramirez F., Lynch J.R., Sykes B.,
RA Panglilan T., Bonadio J.;
RT "Genomic organization of the sequence coding for fibrillin, the
RT defective gene product in Marfan syndrome.";
RL Hum. Mol. Genet. 2:961-968(1993).
RN [2]
RP SEQUENCE OF 1-932 FROM N.A.
RC TISSUE=Placenta; and Fibroblast;
RX MEDLINE=94010947; PubMed=7691719;
RA Corson G.M., Chalberg S.C., Dietz H.C., Charbonneau N.L., Sakai L.Y.;
RT "Fibrillin binds calcium and is coded by cDNAs that reveal a
RT multidomain structure and alternatively spliced exons at the 5'
RT end.";
RL Genomics 17:476-484(1993).
RN [3]
RP SEQUENCE OF 899-2871 FROM N.A.
RX MEDLINE=91304568; PubMed=1852207;
RA Madlen C.L., Corson G.M., Maddox B.K., Glanville R.W., Sakai L.Y.;
RT "Partial sequence of a candidate gene for the Marfan syndrome.";
RL Nature 352:334-337(1991).
RN [4]
RP SEQUENCE OF 813-1313 FROM N.A.
RX MEDLINE=91304567; PubMed=1852206;

RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
RA Tsipouras P., Ramirez F., Hollister D.W.;
RT "Linkage of Marfan syndrome and a phenotypically related disorder to
RT two different fibrillin genes."; *Nature* 352:330-334(1991).
RL Nature 352:330-334(1991).
RN [5]
RP CHARACTERIZATION, PubMed-1860873;
RX MEDLINE-91317849; PubMed-1860873;
RA Sakai L.Y., Keene D.R., Ghanville R.W., Bachinger H.P.;
RT "Purification and partial characterization of fibrillin, a cysteine-
RT rich structural component of connective tissue microfibrils."; *J. Biol. Chem.* 266:14763-14770(1991).
RL J. Biol. Chem. 266:14763-14770(1991).
RN [6]
RP STRUCTURE BY NMR OF 2054-2125.
RX MEDLINE-98031893; PubMed-9362480;
RA Yuan X., Downing A.K., Knott V., Handford P.A.;
RT "Solution structure of the transforming growth factor beta-binding
RT protein-like module, a domain associated with matrix fibrils."; *EMBO J.* 16:6659-6666(1997).
RL EMBO J. 16:6659-6666(1997).
RN [7]
RP STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE-96144829; PubMed-8568869;
RA Knott V., Downing A.K., Cardy C.M., Handford P.A.;
RT "Calcium binding properties of an epidermal growth factor-like domain
RT pair from human fibrillin-1."; *J. Mol. Biol.* 255:22-27(1996).
RL J. Mol. Biol. 255:22-27(1996).
RN [8]
RP STRUCTURE BY NMR OF 2124-2205.
RX MEDLINE-96222301; PubMed-865394;
RA Downing A.K., Knott V., Werner J.M., Cardy C.M., Campbell I.D.,
RA Handford P.A.;
RT "Solution structure of a pair of calcium-binding epidermal growth
RT factor-like domains: implications for the Marfan syndrome and other
RT genetic disorders."; *Cell* 85:597-605(1996).
RL Cell 85:597-605(1996).
RN [9]
RP REVIEW ON MFS VARIANTS.
RX MEDLINE-96174615; PubMed-8594563;
RA Colido G., Beroud C., Soussi T., Junten C., Boileau C.;
RT "Software and database for the analysis of mutations in the human
RT FBN1 gene."; *Nucleic Acids Res.* 24:137-141(1996).
RL Nucleic Acids Res. 24:137-141(1996).
RN [10]
RP REVIEW ON MFS VARIANTS.
RX MEDLINE-97169383; PubMed-9016526;
RA Colido-Beroud G., Beroud C., Ades L., Black C., Boxer M., Brock D.J.,
RA Godfrey M., Hayward C., Karttunen L., Mllewicz D., Peltonen L.,
RA Richards R.I., Wang W., Junten C., Boileau C.;
RT "Marfan Database (second edition): software and database for the
RT analysis of mutations in the human FBN1 gene."; *Nucleic Acids Res.* 25:147-150(1997).
RL Nucleic Acids Res. 25:147-150(1997).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE-98062175; PubMed-9401003;
RA Hayward C., Brock D.J.H.;
RT "Fibrillin-1 mutations in Marfan syndrome and other type-1
RT fibrillinopathies."; *Hum. Mutat.* 10:415-423(1997).
RL Hum. Mutat. 10:415-423(1997).
RN [12]
RP VARIANT MFS PRO-1137.
RX MEDLINE-91304569; PubMed-1852208;
RA Dietz H.C., Cutting G.R., Pyeritz R.E., Maslen C.L., Sakai L.Y.,
RA Corson G.M., Puffenberger E.G., Hamosh A., Nanthakumar E.J.,
RA Curstlin S.M., Stetten G., Meyers D.A., Francomano C.A.;
RT "Marfan syndrome caused by a recurrent de novo missense mutation in
RT the fibrillin gene."; *Nature* 352:337-339(1991).
RL Nature 352:337-339(1991).
RN [13]
RP VARIANTS MFS SER-1249; ARC-1663; SER-2221 AND SER-2307.
RX MEDLINE-93250834; PubMed-1301946;
RA Dietz H.C., Saraiya J.M., Pyeritz R.E., Cutting G.R., Francomano C.A.;
RT "Clustering of fibrillin (FBN1) missense mutations in Marfan syndrome
RT patients at cysteine residues in EGF-like domains."; *Hum. Mutat.* 1:366-374(1992).
RL Hum. Mutat. 1:366-374(1992).
RN [14]
RP VARIANT MFS SER-2307.
RX MEDLINE-92235290; PubMed-1569206;
RA Dietz H.C., Pyeritz R.E., Puffenberger E.G., Kendior R.J., Jr.,
RA Corson G.M., Maslen C.L., Sakai L.Y., Francomano C.A., Cutting G.R.;
RT "Marfan phenotype variability in a family segregating a missense
RT mutation in the epidermal growth factor-like motif of the fibrillin
RT gene."; *J. Clin. Invest.* 89:1674-1680(1992).
RL J. Clin. Invest. 89:1674-1680(1992).
RN [15]
RP VARIANTS MFS ILE-548 AND ALA-723.
RX MEDLINE-94010946; PubMed-8406497;
RA Dietz H.C., McIntosh I., Sakai L.Y., Corson G.M., Chalberg S.C.,
RA Pyeritz R.E., Francomano C.A.;
RT "Four novel FBN1 mutations: significance for mutant transcript level
RT and EGF-like domain calcium binding in the pathogenesis of Marfan
RT syndrome."; *Hum. Mol. Genet.* 17:468-475(1993).
RL Hum. Mol. Genet. 17:468-475(1993).
RN [16]
RP VARIANTS MFS SER-2144.
RX MEDLINE-93278402; PubMed-8504310;
RA Hewett D.R., Lynch J.R., Smith R., Sykes B.C.;
RT "A novel fibrillin mutation in the Marfan syndrome which could
RT disrupt calcium binding of the epidermal growth factor-like module.";
RL Hum. Mol. Genet. 2:475-477(1993).
RN [17]
RP VARIANTS MFS R-862; Y-1117; P-1137 AND F-1589, AND VARIANT A-1148.
RX MEDLINE-94108431; PubMed-8281141;
RA Tyan K., Comeau K., Pearson M., Wilgenbus P., Levitt D., Gasner C.,
RA Berg M.A., Miller D.C., Francke U.;
RT "Mutation screening of complete fibrillin-1 coding sequence: report
RT of five new mutations, including two in 8-cysteine domains."; *Hum. Mol. Genet.* 2:1813-1821(1993).
RL Hum. Mol. Genet. 2:1813-1821(1993).
RN [18]
RP VARIANTS MFS GLY-217 AND ARG-2627.
RX MEDLINE-95067970; PubMed-797366;
RA Karttunen L., Raghunath M., Loenqvist L., Peltonen L.;
RT "A compound-heterozygous Marfan patient: two defective fibrillin
RT alleles result in a lethal phenotype."; *Am. J. Hum. Genet.* 55:1083-1091(1994).
RL Am. J. Hum. Genet. 55:1083-1091(1994).
RN [19]
RP VARIANT EL LYS-2447.
RX MEDLINE-94245249; PubMed-8186302;
RA Longqvist L., Child A., Kainulainen K., Davidson R., Puhakka L.,
RA Peltonen L.;
RT "A novel mutation of the fibrillin gene causing ectopia lentis."; *Genomics* 19:573-576(1994).
RL Genomics 19:573-576(1994).
RN [20]
RP VARIANT MFS CYS-627.
RX MEDLINE-94272487; PubMed-8004112;
RA Hayward C., Rae A.L., Porteous M.E.M., Logie L.J., Brock L.J.;
RT "Two novel mutations and a neutral polymorphism in EGF-like domains
RT of the fibrillin gene (FBN1): SSCP screening of exons 15-21 in Marfan
RT syndrome patients."; *Hum. Mol. Genet.* 3:373-375(1994).
RL Hum. Mol. Genet. 3:373-375(1994).
RN [21]
RP VARIANT MFS CYS-122.
RX MEDLINE-94314977; PubMed-8040326;
RA Stahl-Hallengren C., Urkonen T., Kainulainen K., Kristofersson U.,
RA Saxne T., Tornqvist K., Peltonen L.;
RT "An extra cysteine in one of the non-calcium-binding epidermal growth
RT factor-like motifs of the FBN1 polypeptide is connected to a novel
RT variant of Marfan syndrome."; *J. Clin. Invest.* 94:709-713(1994).
RL J. Clin. Invest. 94:709-713(1994).
RN [22]
RP VARIANT MFS TYR-1223.
RX MEDLINE-94351682; PubMed-8071963;
RA Hewett D.R., Lynch J.R., Child A., Sykes B.C.;
RT "A new missense mutation of fibrillin in a patient with Marfan
RT syndrome."; *J. Med. Genet.* 31:338-339(1994).
RL J. Med. Genet. 31:338-339(1994).
RN [23]

RP VARIANT MFS HIS-1170.
 RA MEDLINE=95174777; PubMed=7870075;
 RA Hayward C., Porteous M.E.M., Brock D.J.H.;
 RT "A novel mutation in the fibrillin gene (FBN1) in familial
 arachnodactyly.";
 RL Mol. Cell. Probes 8:325-327(1994).
 RN [24]
 RP VARIANTS MFS G-217; N-1023; R-1074; Y-1242; R-1513; E-2127; W-2151;
 RP K-2447 AND R-2511.
 RX MEDLINE=94184368; PubMed=8136837;
 RA Kainulainen K., Karttunen L., Punakka L., Sakai L., Peltonen L.;
 RT "Mutations in the fibrillin gene responsible for dominant ectopia
 RT lentis and neonatal Marfan syndrome.";
 RL Nat. Genet. 6:64-69(1994).
 RN [25]
 RP VARIANT SER-1127.

Query Match 6.3%; Score 497.5; DB 1; Length 2871;
 Best Local Similarity 21.7%; Pred. No. 8e-27;
 Matches 261; Conservative 121; Mismatches 356; Indels 465; Gaps 70;

QY 189 POCRTITTRACWLAS-LAHNAK--PARG-----EVMKCALGTASV 226
 DB 442 PRVLPRVNTDVCQLVRYLCQNGRCIPTPGYSRCECNKGFOLDLRECEIDVECEKKNPCAG 501
 QY 227 WDGVNGTG--TCQCGLGNGTACETCTEGKYGIHCDQ-ACSCVHRCGSGPLDGS--CD 281
 DB 502 GECINNGSYTCQCRAGYOSTL--TRTECRDIDECLONGRICNNRCIN--TDSFHCY 556
 QY 282 CDVGMR---GVKCDMETITDNC-----NGCHTSANCLDPDGKASCACAGFR--G 328
 DB 557 CNAGFHVTRDGNCE--DMDECSLRNMCLNGMC-----INEGSKYCKCKPFOLAS 606
 QY 329 NGTVCSTAINACET--SNGGC-----346
 DB 607 DGRYCKDINECECTPGICMNGRCVNTGSGYRCEPPLAVGLDGRVCVDTNMRSTCYGK 666
 QY 347 -----STRADC-----KRTTPGNRY-- 361
 DB 667 RGCCIKPLFGAVTSECCCASTEVAFGPCPCPAONSAEQALSSGPGMTSAGSDINE 726
 QY 362 -----CVCKAGYTG--GIVCLEINPCLENHNGCDRNACCTQTG 398
 DB 727 CALDPDLCPNGICENLRGTAKYCICNSGYEVDSTGKNVCVDINECVLSLDC-NGCCRN- 784
 QY 339 PNOAVCNCLPR---YTGAGKVCSLINVCILTN--NGGC--SPPAFCNTEBODQICCTCP 450
 DB 785 PGSEFVCTC-PRGFIYKPLKTCEDIDECSSPCINGVCNKSPPSF-----ICECSS 834
 QY 451 DYTGD--GIVCRGSYGLPLKPNPSYQFQLOEHAVALAGPGEFTVFAPLSSSFNHP 508
 DB 835 ESTIDPRTICETIKGT-----852
 QY 509 RIKMDQGLMSQYLRYHVHVCQQLLD-NLKVTTSATLQ-----GEPVSISV 556
 DB 853 -----CMQTVIDGRCEININGATLKSCSSGGAAMGSPCTL-C 890
 QY 557 SODTVFINNEAKVLS--DI-----ISTNGVIHVIDLLSPKNILITPK 598
 DB 891 QVDPICGKGYSRIKGTQCEDIDECEVFPYGVCKNGLCVNTRSF--KCOCPSGKMTL--- 943
 QY 599 DALGRVIONLTVAANHGTYKFSKLIODSGLSVI--TDSIHTPVVFMPTDKA----- 650
 DB 944 DATGRICIDIL-----ETCFLEDEDECTLLPIAGRRHMACCCSVCAAMTECECECP 997
 QY 651 -----LEALPEQODLEFNODNKKLKSYLEFHVIRDSKALASDLPRASAKMTLOGSEL 704
 DB 998 MRNTPVEELCPRGFGA---TKEITNGKPEFFKIDINECKMIPSLCTHGKCRNTI--GSF 1051
 QY 705 SVRCGTG-----SDIGELFLNEQMKRFIHHGLF-----DGVAVG----- 740
 DB 1052 KCRCDSGALDSEBRNCTIDECRISPDLC---GRGQCVNTPGFECKCDEGYESGFMM 1108

QY 741 -----IDCLLNPTL--GG-----RCDTFTTFDIPGEGSCIFTPKCPKLSK--PK 782
 DB 1109 KNOMIDECQDPRLLCRGVCHNTEGSRCECPRHQSLSPNISACIDINEELSHLCPN 1168
 QY 783 G-----VKKRCIYNP-----LP--FRNVE-----GCONLCTVVIQTTPRC-CHGYF- 820
 DB 1169 GRCVNLICKYOCACNPGHSTPDLRFVCDIDECSSIMNGCEFFCTNSGSEYSCQPF 1228
 QY 821 -MPDQACRG--GPDTP--CNRMCRDL---YTPMOGCICHGFNCTACELCHHGRFG 871
 DB 1229 LMPDRSCTDIDECEDNPNIDCG-GQCTNIPGEY---RCLCYDEFMASE----- 1273
 QY 872 PDCGRSCSEHGQCD-----EGITSGGECICEGTMTAASCDPTPAVFAVC--T 917
 DB 1274 ---DMKTCVDVNECDNLNICSSTGCENYTKSFLCHCDMGYSKKKRGCGDINCELGA 1330
 QY 918 PACSVHANCETENN--CVCNLNVEBGDITTCVVDPCFCKONNGSCARVAKCSOKGTQVS 972
 DB 1331 HNCGRHAVCT--NFGSEFKSCSPGIDGJICTDLDECSNCTHMCOSHADCKNTMGSYR 1388
 QY 973 CSCKKGYGDCYSCIEIDPCADGV-----NGCCHENATCRMTGPKHKCEKSHYV--GDG 1026
 DB 1389 CLCKRGYTGDDFTCTDLDCESENLMLCNGQC-----LNAFGYRCECDMGFVPSADG 1441
 QY 1027 VDCEPEQLPLDRCLODN---GQCHPDASCADLYPDFTVGVFHLRSPLGQYKILTFDKAK 1082
 DB 1442 KACE-----DIDECSLPNICVGTCH-----NLPLGRCECEIG-YEL--DRSG 1482
 QY 1083 EAC 1085
 DB 1483 GNC 1485

RESULT 4
 FBN2_HUMAN
 ID FBN2_HUMAN STANDARD; PRT; 2911 AA.
 AC P35556;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fibrillin 2 precursor.
 GN FBN2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP MEDLINE=8120105;
 RX SEQUENCE FROM N.A.
 RA Zhang H., Apfelroth S.D., Hu W., Davis E.C., Sanguinetti C.,
 RA Bonadio J., Mecham R.P., Ramirez F.;
 RT "Structure and expression of fibrillin-2, a novel microfibrillar
 RT component preferentially located in elastic matrices.";
 RL J. Cell Biol. 124:855-863(1994).
 RN [2]
 RP SEQUENCE OF 752-1505 FROM N.A.
 RX MEDLINE=91304567; PubMed=1852206;
 RA Lee B., Godfrey M., Vitale E., Hori H., Mattei M.-G., Sarfarazi M.,
 RA Tsipouras P., Ramirez F., Hollister D.;
 RT "Linkage of Marfan syndrome and a phenotypically related disorder to
 RT two different fibrillin genes.";
 RL Nature 352:330-334(1991).
 RN [3]
 RP VARIANTS CCA TYR-1252 AND SER-1433, AND VARIANT ILE-964.
 RX MEDLINE=96083599; PubMed=7493032;
 RA Punam E.A., Zhang H., Ramirez F., Mlejnick D.M.;
 RT "Fibrillin-2 (FBN2) mutations result in the Marfan-like disorder,
 RT congenital contractual arachnodactyly.";
 RL Nat. Genet. 11:456-458(1995).
 RN [4]
 RP VARIANTS CCA HIS-1114.
 RX MEDLINE=98407789; PubMed=9737771;
 RA Babcock D., Gasner C., Francke U., Maslen C.;

DR PIR; A35844; A35844.
DR HSSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF000023; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_Ca; 23.
DR SMART; SM00001; EGF_Like; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 23.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01165; EGF_2; 29.
DR PROSITE; PS01187; EGF_CA; 21.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 2524
FT DOMAIN 20 1728
FT TRANSMEM 1729 1750
FT DOMAIN 1751 2524
FT DOMAIN 20 58
FT DOMAIN 58 99
FT DOMAIN 102 140
FT DOMAIN 141 177
FT DOMAIN 179 215
FT DOMAIN 217 254
FT DOMAIN 256 292
FT DOMAIN 294 332
FT DOMAIN 334 370
FT DOMAIN 371 409
FT DOMAIN 411 449
FT DOMAIN 451 487
FT DOMAIN 489 525
FT DOMAIN 527 563
FT DOMAIN 565 600
FT DOMAIN 602 638
FT DOMAIN 640 675
FT DOMAIN 677 713
FT DOMAIN 715 750
FT DOMAIN 752 788
FT DOMAIN 790 826
FT DOMAIN 828 866
FT DOMAIN 868 904
FT DOMAIN 906 942
FT DOMAIN 944 980
FT DOMAIN 982 1018
FT DOMAIN 1020 1056
FT DOMAIN 1058 1094
FT DOMAIN 1096 1142
FT DOMAIN 1144 1180
FT DOMAIN 1182 1218
FT DOMAIN 1220 1264
FT DOMAIN 1266 1304
FT DOMAIN 1306 1346
FT DOMAIN 1347 1383
FT DOMAIN 1386 1424
FT DOMAIN 1441 1478
FT REPEAT 1479 1520
FT REPEAT 1521 1560
FT REPEAT 1561 1600
FT REPEAT 1601 1640
FT REPEAT 1641 1680
FT REPEAT 1681 1720
FT REPEAT 1721 1760
FT REPEAT 1761 1800
FT REPEAT 1801 1840
FT REPEAT 1841 1880
FT REPEAT 1881 1920
FT REPEAT 1921 1960
FT REPEAT 1961 2000
FT REPEAT 2001 2040
FT REPEAT 2041 2080
FT REPEAT 2081 2120
FT REPEAT 2121 2160
FT REPEAT 2161 2200
FT REPEAT 2201 2240
FT REPEAT 2241 2280
FT REPEAT 2281 2320
FT REPEAT 2321 2360
FT REPEAT 2361 2400
FT REPEAT 2401 2440
FT REPEAT 2441 2480
FT REPEAT 2481 2520
FT REPEAT 2521 2560
FT REPEAT 2561 2600
FT REPEAT 2601 2640
FT REPEAT 2641 2680
FT REPEAT 2681 2720
FT REPEAT 2721 2760
FT REPEAT 2761 2800
FT REPEAT 2801 2840
FT REPEAT 2841 2880
FT REPEAT 2881 2920
FT REPEAT 2921 2960
FT REPEAT 2961 3000
FT REPEAT 3001 3040
FT REPEAT 3041 3080
FT REPEAT 3081 3120
FT REPEAT 3121 3160
FT REPEAT 3161 3200
FT REPEAT 3201 3240
FT REPEAT 3241 3280
FT REPEAT 3281 3320
FT REPEAT 3321 3360
FT REPEAT 3361 3400
FT REPEAT 3401 3440
FT REPEAT 3441 3480
FT REPEAT 3481 3520
FT REPEAT 3521 3560
FT REPEAT 3561 3600
FT REPEAT 3601 3640
FT REPEAT 3641 3680
FT REPEAT 3681 3720
FT REPEAT 3721 3760
FT REPEAT 3761 3800
FT REPEAT 3801 3840
FT REPEAT 3841 3880
FT REPEAT 3881 3920
FT REPEAT 3921 3960
FT REPEAT 3961 4000
FT REPEAT 4001 4040
FT REPEAT 4041 4080
FT REPEAT 4081 4120
FT REPEAT 4121 4160
FT REPEAT 4161 4200
FT REPEAT 4201 4240
FT REPEAT 4241 4280
FT REPEAT 4281 4320
FT REPEAT 4321 4360
FT REPEAT 4361 4400
FT REPEAT 4401 4440
FT REPEAT 4441 4480
FT REPEAT 4481 4520
FT REPEAT 4521 4560
FT REPEAT 4561 4600
FT REPEAT 4601 4640
FT REPEAT 4641 4680
FT REPEAT 4681 4720
FT REPEAT 4721 4760
FT REPEAT 4761 4800
FT REPEAT 4801 4840
FT REPEAT 4841 4880
FT REPEAT 4881 4920
FT REPEAT 4921 4960
FT REPEAT 4961 5000
FT REPEAT 5001 5040
FT REPEAT 5041 5080
FT REPEAT 5081 5120
FT REPEAT 5121 5160
FT REPEAT 5161 5200
FT REPEAT 5201 5240
FT REPEAT 5241 5280
FT REPEAT 5281 5320
FT REPEAT 5321 5360
FT REPEAT 5361 5400
FT REPEAT 5401 5440
FT REPEAT 5441 5480
FT REPEAT 5481 5520
FT REPEAT 5521 5560
FT REPEAT 5561 5600
FT REPEAT 5601 5640
FT REPEAT 5641 5680
FT REPEAT 5681 5720
FT REPEAT 5721 5760
FT REPEAT 5761 5800
FT REPEAT 5801 5840
FT REPEAT 5841 5880
FT REPEAT 5881 5920
FT REPEAT 5921 5960
FT REPEAT 5961 6000
FT REPEAT 6001 6040
FT REPEAT 6041 6080
FT REPEAT 6081 6120
FT REPEAT 6121 6160
FT REPEAT 6161 6200
FT REPEAT 6201 6240
FT REPEAT 6241 6280
FT REPEAT 6281 6320
FT REPEAT 6321 6360
FT REPEAT 6361 6400
FT REPEAT 6401 6440
FT REPEAT 6441 6480
FT REPEAT 6481 6520
FT REPEAT 6521 6560
FT REPEAT 6561 6600
FT REPEAT 6601 6640
FT REPEAT 6641 6680
FT REPEAT 6681 6720
FT REPEAT 6721 6760
FT REPEAT 6761 6800
FT REPEAT 6801 6840
FT REPEAT 6841 6880
FT REPEAT 6881 6920
FT REPEAT 6921 6960
FT REPEAT 6961 7000
FT REPEAT 7001 7040
FT REPEAT 7041 7080
FT REPEAT 7081 7120
FT REPEAT 7121 7160
FT REPEAT 7161 7200
FT REPEAT 7201 7240
FT REPEAT 7241 7280
FT REPEAT 7281 7320
FT REPEAT 7321 7360
FT REPEAT 7361 7400
FT REPEAT 7401 7440
FT REPEAT 7441 7480
FT REPEAT 7481 7520
FT REPEAT 7521 7560
FT REPEAT 7561 7600
FT REPEAT 7601 7640
FT REPEAT 7641 7680
FT REPEAT 7681 7720
FT REPEAT 7721 7760
FT REPEAT 7761 7800
FT REPEAT 7801 7840
FT REPEAT 7841 7880
FT REPEAT 7881 7920
FT REPEAT 7921 7960
FT REPEAT 7961 8000
FT REPEAT 8001 8040
FT REPEAT 8041 8080
FT REPEAT 8081 8120
FT REPEAT 8121 8160
FT REPEAT 8161 8200
FT REPEAT 8201 8240
FT REPEAT 8241 8280
FT REPEAT 8281 8320
FT REPEAT 8321 8360
FT REPEAT 8361 8400
FT REPEAT 8401 8440
FT REPEAT 8441 8480
FT REPEAT 8481 8520
FT REPEAT 8521 8560
FT REPEAT 8561 8600
FT REPEAT 8601 8640
FT REPEAT 8641 8680
FT REPEAT 8681 8720
FT REPEAT 8721 8760
FT REPEAT 8761 8800
FT REPEAT 8801 8840
FT REPEAT 8841 8880
FT REPEAT 8881 8920
FT REPEAT 8921 8960
FT REPEAT 8961 9000
FT REPEAT 9001 9040
FT REPEAT 9041 9080
FT REPEAT 9081 9120
FT REPEAT 9121 9160
FT REPEAT 9161 9200
FT REPEAT 9201 9240
FT REPEAT 9241 9280
FT REPEAT 9281 9320
FT REPEAT 9321 9360
FT REPEAT 9361 9400
FT REPEAT 9401 9440
FT REPEAT 9441 9480
FT REPEAT 9481 9520
FT REPEAT 9521 9560
FT REPEAT 9561 9600
FT REPEAT 9601 9640
FT REPEAT 9641 9680
FT REPEAT 9681 9720
FT REPEAT 9721 9760
FT REPEAT 9761 9800
FT REPEAT 9801 9840
FT REPEAT 9841 9880
FT REPEAT 9881 9920
FT REPEAT 9921 9960
FT REPEAT 9961 10000
FT REPEAT 10001 10040
FT REPEAT 10041 10080
FT REPEAT 10081 10120
FT REPEAT 10121 10160
FT REPEAT 10161 10200
FT REPEAT 10201 10240
FT REPEAT 10241 10280
FT REPEAT 10281 10320
FT REPEAT 10321 10360
FT REPEAT 10361 10400
FT REPEAT 10401 10440
FT REPEAT 10441 10480
FT REPEAT 10481 10520
FT REPEAT 10521 10560
FT REPEAT 10561 10600
FT REPEAT 10601 10640
FT REPEAT 10641 10680
FT REPEAT 10681 10720
FT REPEAT 10721 10760
FT REPEAT 10761 10800
FT REPEAT 10801 10840
FT REPEAT 10841 10880
FT REPEAT 10881 10920
FT REPEAT 10921 10960
FT REPEAT 10961 11000
FT REPEAT 11001 11040
FT REPEAT 11041 11080
FT REPEAT 11081 11120
FT REPEAT 11121 11160
FT REPEAT 11161 11200
FT REPEAT 11201 11240
FT REPEAT 11241 11280
FT REPEAT 11281 11320
FT REPEAT 11321 11360
FT REPEAT 11361 11400
FT REPEAT 11401 11440
FT REPEAT 11441 11480
FT REPEAT 11481 11520
FT REPEAT 11521 11560
FT REPEAT 11561 11600
FT REPEAT 11601 11640
FT REPEAT 11641 11680
FT REPEAT 11681 11720
FT REPEAT 11721 11760
FT REPEAT 11761 11800
FT REPEAT 11801 11840
FT REPEAT 11841 11880
FT REPEAT 11881 11920
FT REPEAT 11921 11960
FT REPEAT 11961 12000
FT REPEAT 12001 12040
FT REPEAT 12041 12080
FT REPEAT 12081 12120
FT REPEAT 12121 12160
FT REPEAT 12161 12200
FT REPEAT 12201 12240
FT REPEAT 12241 12280
FT REPEAT 12281 12320
FT REPEAT 12321 12360
FT REPEAT 12361 12400
FT REPEAT 12401 12440
FT REPEAT 12441 12480
FT REPEAT 12481 12520
FT REPEAT 12521 12560
FT REPEAT 12561 12600
FT REPEAT 12601 12640
FT REPEAT 12641 12680
FT REPEAT 12681 12720
FT REPEAT 12721 12760
FT REPEAT 12761 12800
FT REPEAT 12801 12840
FT REPEAT 12841 12880
FT REPEAT 12881 12920
FT REPEAT 12921 12960
FT REPEAT 12961 13000
FT REPEAT 13001 13040
FT REPEAT 13041 13080
FT REPEAT 13081 13120
FT REPEAT 13121 13160
FT REPEAT 13161 13200
FT REPEAT 13201 13240
FT REPEAT 13241 13280
FT REPEAT 13281 13320
FT REPEAT 13321 13360
FT REPEAT 13361 13400
FT REPEAT 13401 13440
FT REPEAT 13441 13480
FT REPEAT 13481 13520
FT REPEAT 13521 13560
FT REPEAT 13561 13600
FT REPEAT 13601 13640
FT REPEAT 13641 13680
FT REPEAT 13681 13720
FT REPEAT 13721 13760
FT REPEAT 13761 13800
FT REPEAT 13801 13840
FT REPEAT 13841 13880
FT REPEAT 13881 13920
FT REPEAT 13921 13960
FT REPEAT 13961 14000
FT REPEAT 14001 14040
FT REPEAT 14041 14080
FT REPEAT 14081 14120
FT REPEAT 14121 14160
FT REPEAT 14161 14200
FT REPEAT 14201 14240
FT REPEAT 14241 14280
FT REPEAT 14281 14320
FT REPEAT 14321 14360
FT REPEAT 14361 14400
FT REPEAT 14401 14440
FT REPEAT 14441 14480
FT REPEAT 14481 14520
FT REPEAT 14521 14560
FT REPEAT 14561 14600
FT REPEAT 14601 14640
FT REPEAT 14641 14680
FT REPEAT 14681 14720
FT REPEAT 14721 14760
FT REPEAT 14761 14800
FT REPEAT 14801 14840
FT REPEAT 14841 14880
FT REPEAT 14881 14920
FT REPEAT 14921 14960
FT REPEAT 14961 15000
FT REPEAT 15001 15040
FT REPEAT 15041 15080
FT REPEAT 15081 15120
FT REPEAT 15121 15160
FT REPEAT 15161 15200
FT REPEAT 15201 15240
FT REPEAT 15241 15280
FT REPEAT 15281 15320
FT REPEAT 15321 15360
FT REPEAT 15361 15400
FT REPEAT 15401 15440
FT REPEAT 15441 15480
FT REPEAT 15481 15520
FT REPEAT 15521 15560
FT REPEAT 15561 15600
FT REPEAT 15601 15640
FT REPEAT 15641 15680
FT REPEAT 15681 15720
FT REPEAT 15721 15760
FT REPEAT 15761 15800
FT REPEAT 15801 15840
FT REPEAT 15841 15880
FT REPEAT 15881 15920
FT REPEAT 15921 15960
FT REPEAT 15961 16000
FT REPEAT 16001 16040
FT REPEAT 16041 16080
FT REPEAT 16081 16120
FT REPEAT 16121 16160
FT REPEAT 16161 16200
FT REPEAT 16201 16240
FT REPEAT 16241 16280
FT REPEAT 16281 16320
FT REPEAT 16321 16360
FT REPEAT 16361 16400
FT REPEAT 16401 16440
FT REPEAT 16441 16480
FT REPEAT 16481 16520
FT REPEAT 16521 16560
FT REPEAT 16561 16600
FT REPEAT 16601 16640
FT REPEAT 16641 16680
FT REPEAT 16681 16720
FT REPEAT 16721 16760
FT REPEAT 16761 16800
FT REPEAT 16801 16840
FT REPEAT 16841 16880
FT REPEAT 16881 16920
FT REPEAT 16921 16960
FT REPEAT 16961 17000
FT REPEAT 17001 17040
FT REPEAT 17041 17080
FT REPEAT 17081 17120
FT REPEAT 17121 17160
FT REPEAT 17161 17200
FT REPEAT 17201 17240
FT REPEAT 17241 17280
FT REPEAT 17281 17320
FT REPEAT 17321 17360
FT REPEAT 17361 17400
FT REPEAT 17401 17440
FT REPEAT 17441 17480
FT REPEAT 17481 17520
FT REPEAT 17521 17560
FT REPEAT 17561 17600
FT REPEAT 17601 17640
FT REPEAT 17641 17680
FT REPEAT 17681 17720
FT REPEAT 17721 17760
FT REPEAT 17761 17800
FT REPEAT 17801 17840
FT REPEAT 17841 17880
FT REPEAT 17881 17920
FT REPEAT 17921 17960
FT REPEAT 17961 18000
FT REPEAT 18001 18040
FT REPEAT 18041 18080
FT REPEAT 18081 18120
FT REPEAT 18121 18160
FT REPEAT 18161 18200
FT REPEAT 18201 18240
FT REPEAT 18241 18280
FT REPEAT 18281 18320
FT REPEAT 18321 18360
FT REPEAT 18361 18400
FT REPEAT 18401 18440
FT REPEAT 18441 18480
FT REPEAT 18481 18520
FT REPEAT 18521 18560
FT REPEAT 18561 18600
FT REPEAT 18601 18640
FT REPEAT 18641 18680
FT REPEAT 18681 18720
FT REPEAT 18721 18760
FT REPEAT 18761 18800
FT REPEAT 18801 18840
FT REPEAT 18841 18880
FT REPEAT 18881 18920
FT REPEAT 18921 18960
FT REPEAT 18961 19000
FT REPEAT 19001 19040
FT REPEAT 19041 19080
FT REPEAT 19081 19120
FT REPEAT 19121 19160
FT REPEAT 19161 19200
FT REPEAT 19201 19240
FT REPEAT 19241 19280
FT REPEAT 19281 19320
FT REPEAT 19321 19360
FT REPEAT 19361 19400
FT REPEAT 19401 19440
FT REPEAT 19441 19480
FT REPEAT 19481 19520
FT REPEAT 19521 19560
FT REPEAT 19561 19600
FT REPEAT 19601 19640
FT REPEAT 19641 19680
FT REPEAT 19681 19720
FT REPEAT 19721 19760
FT REPEAT 19761 19800
FT REPEAT 19801 19840
FT REPEAT 19841 19880
FT REPEAT 19881 19920
FT REPEAT 19921 19960
FT REPEAT 19961 20000
FT REPEAT 20001 20040
FT REPEAT 20041 20080
FT REPEAT 20081 20120
FT REPEAT 20121 20160
FT REPEAT 20161 20200
FT REPEAT 20201 20240
FT REPEAT 20241 20280
FT REPEAT 20281 20320
FT REPEAT 20321 20360
FT REPEAT 20361 20400
FT REPEAT 20401 20440
FT REPEAT 20441 20480
FT REPEAT 20481 20520
FT REPEAT 20521 20560
FT REPEAT 20561 20600
FT REPEAT 20601 20640
FT REPEAT 20641 20680
FT REPEAT 20681 20720
FT REPEAT 20721 20760
FT REPEAT 20761 20800
FT REPEAT 20801 20840
FT REPEAT 20841 20880
FT REPEAT 20881 20920
FT REPEAT 20921 20960
FT REPEAT 20961 21000
FT REPEAT 21001 21040
FT REPEAT 21041 21080
FT REPEAT 21081 21120
FT REPEAT 21121 21160
FT REPEAT 21161 21200
FT REPEAT 21201 21240
FT REPEAT 21241 21280
FT REPEAT 21281 21320
FT REPEAT 21321 21360
FT REPEAT 21361 21400
FT REPEAT 21401 21440
FT REPEAT 21441 21480
FT REPEAT 21481 21520
FT REPEAT 21521 21560
FT REPEAT 21561 21600
FT REPEAT 21601 21640
FT REPEAT 21641 21680
FT REPEAT 21681 21720
FT REPEAT 21721 21760
FT REPEAT 21761 21800
FT REPEAT 21801 21840
FT REPEAT 21841 21880
FT REPEAT 21881 21920
FT REPEAT 21921 21960
FT REPEAT 21961 22000
FT REPEAT 22001 22040
FT REPEAT 22041 22080
FT REPEAT 22081 22120
FT REPEAT 22121 22160
FT REPEAT 22161 22200
FT REPEAT 22201 22240
FT REPEAT 22241 22280
FT REPEAT 22281 22320
FT REPEAT 22321 22360
FT REPEAT 22361 22400
FT REPEAT 22401 22440
FT REPEAT 22441 22480
FT REPEAT 22481 22520
FT REPEAT 22521 22560
FT REPEAT 22561 22600
FT REPEAT 22601 22640
FT REPEAT 22641 22680
FT REPEAT 22681 22720
FT REPEAT 22721 22760
FT REPEAT 22761 22800
FT REPEAT 22801 22840
FT REPEAT 22841 22880
FT REPEAT 22881 22920
FT REPEAT 22921 22960
FT REPEAT 22961 23000
FT REPEAT 23001 23040
FT REPEAT 23041 23080
FT REPEAT 23081 23120
FT REPEAT 23121 23160
FT REPEAT 23161 23200
FT REPEAT 23201 23240
FT REPEAT 23241 23280
FT REPEAT 23281 23320
FT REPEAT 23321 23360
FT REPEAT 23361 23400
FT REPEAT 23401 23440
FT REPEAT 23441 23480
FT REPEAT 23481 23520
FT REPEAT 23521 23560
FT REPEAT 23561 23600
FT REPEAT 23601 23640
FT REPEAT 23641 23680
FT REPEAT 23681 23720
FT REPEAT 23721 23760
FT REPEAT 23761 23800
FT REPEAT 23801 23840
FT REPEAT 23841 23880
FT REPEAT 23881 23920
FT REPEAT 23921 23960
FT REPEAT 23961 24000
FT REPEAT 24001 24040
FT REPEAT 24041 24080
FT REPEAT 24081 24120
FT REPEAT 24121 24160
FT REPEAT 24161 24200
FT REPEAT 24201 24240
FT REPEAT 24241 24280
FT REPEAT 24281 24320
FT REPEAT 24321 24360
FT REPEAT 24361 24400
FT REPEAT 24401 24440
FT REPEAT 24441 24480
FT REPEAT 24481 24520
FT REPEAT 24521 24560
FT REPEAT 24561 24600
FT REPEAT 24601 24640
FT REPEAT 24641 24680
FT REPEAT 24681 24720
FT REPEAT 24721 24760
FT REPEAT 24761 24800
FT REPEAT 24801 24840
FT REPEAT 24841 24880
FT REPEAT 24881 24920
FT REPEAT 24921 24960
FT REPEAT 24961 25000
FT REPEAT 25001 25040
FT REPEAT 25041 25080
FT REPEAT 25081 25120
FT REPEAT 25121 25160
FT REPEAT 25161 25200
FT REPEAT 25201 25240
FT REPEAT 25241 25280
FT REPEAT 25281 25320
FT REPEAT 25321 25360
FT REPEAT 25361 25400
FT REPEAT 25401 25440
FT REPEAT 25441 25480
FT REPEAT 25481 25520
FT REPEAT 25521 25560
FT REPEAT 25561 25600
FT REPEAT 25601 25640
FT REPEAT 25641 25680
FT REPEAT 25681 25720
FT REPEAT 25721 25760
FT REPEAT 25761 25800
FT REPEAT 25801 25840
FT REPEAT 25841 25880
FT REPEAT 25881 25920
FT REPEAT 25921 25960
FT REPEAT 25961 26000
FT REPEAT 26001 26040
FT REPEAT 26041 26080
FT REPEAT 26081 26120
FT REPEAT 26121 26160
FT REPEAT 26161 26200
FT REPEAT 26201 26240
FT REPEAT 26241 26280
FT REPEAT 26281 26320
FT REPEAT 26321 26360
FT REPEAT 26361 26400
FT REPEAT 26401 26440
FT REPEAT 26441 26480
FT REPEAT 26481 26520
FT REPEAT 26521 26560
FT REPEAT 26561 26600
FT REPEAT 26601 26640
FT REPEAT 26641 26680
FT REPEAT 26681 26720
FT REPEAT 26721 26760
FT REPEAT 26761 26800
FT REPEAT 26801 26840
FT REPEAT 26841 26880
FT REPEAT 26881 26920
FT REPEAT 26921 26960
FT REPEAT 26961 27000
FT REPEAT 27001 27040
FT REPEAT 27041 27080
FT REPEAT 27081 27120
FT REPEAT 27121 27160
FT REPEAT 27161 27200
FT REPEAT 27201 27240
FT REPEAT 27241 27280
FT REPEAT 27281 27320
FT REPEAT 27321 27360
FT REPEAT 27361 27400
FT REPEAT 27401 27440
FT REPEAT 27441 27480
FT REPEAT 27481 27520
FT REPEAT 27521 27560
FT REPEAT 27561 27600
FT REPEAT 27601 27640
FT REPEAT 27641 27680
FT REPEAT 27681 27720
FT REPEAT 27721 27760
FT REPEAT 27761 27800
FT REPEAT 27801 27840
FT REPEAT 27841 27880
FT REPEAT 27881 27920
FT REPEAT 27921 27960
FT REPEAT 27961 28000
FT REPEAT 28001 28040
FT REPEAT 28041 28080
FT REPEAT 28081 28120
FT REPEAT 28121 28160
FT REPEAT 28161 28200
FT REPEAT 28201 28240
FT REPEAT 28241 28280
FT REPEAT 28281 28320
FT REPEAT 28321 28360
FT REPEAT 28361 28400
FT REPEAT 28401 28440
FT REPEAT 28441 28480
FT REPEAT 28481 28520
FT REPEAT 28521 28560
FT REPEAT 28561 28600
FT REPEAT 28601 28640
FT REPEAT 28641 28680
FT REPEAT 28681 28720
FT REPEAT 28721 28760
FT REPEAT 28761 28800
FT RE

Query Match
Best Local Similarity 21.9%; Pred. No. 4,4e-26;
Matches 302; Conservative 131; Mismatches 457; Indels 486; Gaps 93;

```

105 YVNAPIINT-----NVAIDKGVHGLEKYLEIQRNCDNDITIVRGECC 150
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
85 FICHPVGFIDKVCITPVNACVNNPCRNCGTCCLNSYTE-YKRCRCPGWT---GD-- 137
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
151 KCSQOAPCPLKTRKCI-YSIYMGKRSVFICGP----- 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
138 SCQOADC--ASNPCANGCKLPFEIOTYCK-----CPGPHGATCKODINECSQNPCK 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 ---OCVFTITRACMIAS--LAHN-----AKPAPEVKKALGTASVMDGVNGTGTQ 237
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
190 NGGQGINEFSGSYRCTQNRFTGRNCDEPYVPCNSP-----C-----LNG-GTCR 233
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
238 ---CGLGFNGTACE-----TCTEG--KYGIHC-----DQAC-- 263
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
234 QTDSDYSCTCLPFGFSQNEENIDCPSNNCRNGGTCVGVNTVNCQCPDWTGYCTE 293
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 -----SVHGRCSGPLGDSCDQVGMGVKCDMEITDNC-NGTCHTSANCL 312
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
294 DVEDCOLMPNACQNGCHNTYGGYNCVANGMTGEDECSNI--DDCANACHSGATC-- 349
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
313 DPGDKAS--CKAAGFRNGTGTCTAINACTSSNGGCTKADCKRTP--GNRVCACAGY 368
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
350 -HBRVASFYCECPHG--RTGLLCHLDNAC--ISNPCNEGSNCD--TNPVNGKAITCPCGY 403
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
369 TG-----DGIVC-LEINPCLEHGGCGRNAECTOT--GRNCAVNCCLKRYTDGKVCSL-I 420
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
404 TGPACNNVDVDECSLGANPC--EHGG-----RCTNTLSFO--CNCPOGYAGPR--CEIDV 452
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
421 NVCLTNNGGCSPEAFCVYTEODORICTCKPDYTGDIYCRGSIYIGELPKNP----- 471
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
453 NECLSN--PCQNDSTC-LDQIGERQICMPCY--EGLXCFENI--DECASNCLNMGKCID 506
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
472 -----STSYQFFQLOEHAHVR-----LAGPGFTVFAPLSSSFNH----- 506
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
507 KINEFRCDCTPFGSGNLCOHDFDECTSTPCKNGAKCLDGRNSYTCQCTEGFTGRHCBODI 566
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
507 EPRIKMDQOGLMSQVLRHYHVGQ-----QLLDNLKVTTSATTTGGEFVS-----I 554
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
567 NECLPDCHYGTCKDGLATFTCLCRPYTGRLCNDINECLSKPCLNNGGQCTDBENGYIC 626
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
555 SVSODVFINNEAKV--LSSDIISTNGVIHYID-----KLSPKLLI----- 595
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
627 TCRPGTTGVNCEIKIDOCASLONCGKCIDKIDYEECTCEGYTGKLC--NININEGDS 683
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
596 -----TPPDALGRVLQNTTVAANHGYTKFSKLIQDSGLSVITD-----STHT 639
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
684 NPCRNGSTCKDDI-----NGFTVCPCPDGY-----HDHMLSEVNCSNPNCLHGACHD 731
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
640 PVYVF-----WPTDKALEALPREODDLEFNQDNKRLKLSYLFHYIRSKALASLDLPR 692
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
732 GVGNGYCKDCERAGW-----SGSNCD-----INNCECSNPMCN 763
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

693 SASMKTLQSGSELSVRCNG-----SDIGELFLNEMQCRFIHRGLLEFDVGAAYGIDCLL 745
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
764 GGTCKDMHTGAICT--CKAGFSGPCMOTNINCCSSNP--C--LNNGTCTIDVAGYKRCNML 818
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
746 -----MNPVL-----GGRC--DREFTREDI--PG-----ECGSCIFPP-- 773
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
819 PYTGATCEAVLAPCAGSPCKNGCRCKESEDEFTSCCECPGMQGTCEIDMNECVNRCR 878
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
774 -----KCPLRKSPKGVKKKCIYNLPFRNVEGQNLCTVYIOTPRCH-- 817
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
879 NGATCQNTNGSYKC--NCKPGYTGANC-----EMDIDDC-----PNCCHNGS 920
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
818 ---GVPMPCOACAPCG--PDTPCNNRMCRLD---YTPMGQCLCHGFN 858
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
921 CSDGLNMFCH--CAGFRGPKCEDINDCASNPCKNANCIDCVNST-----CTCQGFSS 975
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
859 GTACELCMHGRFPDQPRSCSEHGQCDDEGITSGEICETGWTAAACDPPTAVFAVCTP 918
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
976 GHCE-----SNTPDCYESSCFNGTCTIDGI-NFTCQCPGFTGYSYQHD--INECDSK 1027
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
919 ACSVHATCTEN--NTCVCLNLYEGDITC--TVYDFCKQNNNGGCAKAKAGSKGTQVSCS 974
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1028 PCLNGTCTQDSYGTGYKCTCPDGY--TGLNCONLVRWC--DSSPCKNGSKCWQTNMFYRCE 1083
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
975 CKKGKYG-----DGYSCIEIDPCADGVN--GGCHEHATCRMTGQKHKCECKSHYVGDV 1027
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1084 CKSGMTGYCVCPBSVSC--EVAKAQGVDIYHLCRNSCMCVTGT--NTHFCRQAGTYGS-- 1139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1028 DCEPEQPLDRLCLDQNGOCHPDASCADLYFQDITTVGFHLRSLPGOYKLPFDKAKKACAK 1087
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1140 YCEBQ-----VDEC--SPNQCQNGATCTDY-----LGQY----- 1166
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1088 EAATLATYNNOLSTYAKAK---YHLCASG-----WLESRAVAPPTTYSQKCGANVVGIVD 1139
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1167 SCECVAGYHGVNCSSEIINECLSHPCQNGTCTIDLINITYKSCSPRGTOGVHCEIIVNDCTP 1226
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1140 YGSRANKSEMDVF-----CY-----RMKDVCCTCKAGYVGGFSGSNLLQVL 1183
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1227 F-----YDSFTLEPKCFNNCKCIDRVGYNICICPPRGVGE--KCEGDVNECL 1271
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 6
NOTC DROME STANDARD: PRT: 2703 AA.
AC P07207; P04154; O97458; Q9M4T8;
DT 01-NOV-1986 (Rel. 03, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus Notch protein precursor.
GN N OR EG:140611.1 OR EG:163A10.2 OR CG3936.
OS Drosophila melanogaster (Fruit fly).
OC Euarthropoda; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; and Oregon-R; TISSUE=Embryo;
RX MEDLINE=86079539; PubMed=3935325;
RA Wharton K.A., Johansen K.M., Xu T., Artavanis-Tsakonas S.;
RT "Nucleotide sequence from the neurogenic locus notch implies a gene
product that shares homology with proteins containing EGF-like
repeats";
RL Cell 43:567-581(1985).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Canton-S; and Oregon-R; TISSUE=Embryo;
RX MEDLINE=87064624; PubMed=3097517;
RA Kidd S., Kelley M.R., Young M.W.;
RT "Sequence of the notch locus of Drosophila melanogaster: relationship
of the encoded protein to mammalian clotting and growth factors";
ML Mol. Cell. Biol. 6:3094-3108(1986).
RN [3]


```

FT DOMAIN 139 176 EGF-LIKE 3.
FT DOMAIN 177 215 EGF-LIKE 4.
FT DOMAIN 217 253 EGF-LIKE 5. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 255 291 EGF-LIKE 6.
FT DOMAIN 293 329 EGF-LIKE 7. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 331 370 EGF-LIKE 8. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 372 408 EGF-LIKE 9. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 409 447 EGF-LIKE 10.
FT DOMAIN 449 486 EGF-LIKE 11. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 488 524 EGF-LIKE 12. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 526 562 EGF-LIKE 13. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 564 600 EGF-LIKE 14. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 602 637 EGF-LIKE 15. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 639 675 EGF-LIKE 16. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 677 713 EGF-LIKE 17. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 715 751 EGF-LIKE 18. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 753 789 EGF-LIKE 19. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 791 827 EGF-LIKE 20. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 829 865 EGF-LIKE 21. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 867 905 EGF-LIKE 22.
FT DOMAIN 907 944 EGF-LIKE 23. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 946 982 EGF-LIKE 24. CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 984 1020 EGF-LIKE 25.
FT DOMAIN 1022 1058 EGF-LIKE 26. CALCIUM-BINDING (POTENTIAL).

```

Query Match 6.2% Score 484; DB 1; Length 2703;

Best Local Similarity 23.1% Pred. No. 6.7e-26; Mismatches 382; Indels 414; Gaps 75;

```

Matches 269; Conservative 97;
OY 218 MCALCTASYVMDVNGTGTCCGGLGFNGTACE-----TCTEGKXGI--HCDQACS--- 264
DB 359 ICVNGMAGL-DCSNMTDCKQAACFGATCIDGVGSFYQCGTKRGKGLLCHLDDATSNP 417
OY 265 C-VHGRCSGGL-GRGSCDCDVGMRGKVCMDMETTDNCL--GTCHTSAACLLDPDGKASC 320
DB 418 CHADAICDPTSPINGASCAATGYKGVDCSEDI--DECOGSSPCEINGICVMP--GSYRC 474
OY 321 KCAAGFRNGTVC-TAINACFTSNGGCSFKADCKRTTPGNRVCCVCKAGYTG-----DGIV 374
DB 475 MCSQGF--TGPRCFETNINECESH--PCONEGSC-LDDPCTFRVCVCPGFTGQCEIIDE 529
OY 375 CLEINPCLENHGGC-DR-----NAECTGTGP--NOAVCN----- 405
DB 530 C-QSNPCL-NDGTCHDKIKCFKSCALGFTGARCOINIDDCSQSPRRNGICHDSIAGS 587
OY 406 --CLPKRYTGDGVCSL-INVCLTN--NGGSPFAFCNTBEDQRTCTCKPRDTGDIYC 459
DB 588 CECPPGYTGT--CEININDCDSNPCRKGC-----IDVNSFKCLDPPGT--GYIC 636
OY 460 RGSITGELPKNP-----STSQYFPOLOEHAVRELAGPPFTYFAFLSSFNHPR 510
DB 637 OKOI-NECESNPGQFDGHCQDRVGSYYCCQ-----AG-----TSCKNEVNV 678
OY 511 KMDQDGLMSQVLRHYVVCQOOLLNLNKLVTTSATTLGCEPVSISVSQDTVFNNKAYL 570
DB 679 NE-----CHSNPC-----NNCAT--- 691
OY 571 SSDIISTNGVHIVDKLSPKMLI--TPKDALGRVLIQNLTTVAANHGYTFESKLIDS 627
DB 692 -----CIDGINSYKQCQVAGFTGQHCENKVDCEISSPCANNG----- 728
OY 628 GLLSVITDSIHTPVVFWETDKALEALPREQODLFLENOONKRLKLYLFHYIRDSKALA 687
DB 729 ---VCIDOV-----NGKYCECPREFYAHACLSIDVDECA 758
OY 688 SLPSPASAKTIQG-SELSVRCGTG-----SDIGELFLNEOMCFIIRGLFLDVGVAY 739
DB 759 SN-PCVNEBRCEDGINEFICHCPGYTGKRCLEIDDECSNP--CQ--HGCTCYDKLNAF 813
OY 740 GIDCLLAMPITLGRCDTFTTPIGEC--GSCI-----PTPK-CPLKSKRP 782
DB 814 SCQC--MPGYTOKCETNIDCVTNPCGNGGTCIDKVNKGVCVKCVFPGRDCESKMDPC 871

```

```

OY 783 GYKK-----KCI--YNPLFR-----RVNEGCONL----- 805
DB 872 ASNRCKNEAKCTPSSNFLEFDSCTCKLGYRGYRCDIEDCESSLPSCRNASCLANVGSYR 931
OY 806 GTVVIOTPRCHGVEYEMPC-----QACPGSPDPCCNRRMGKCRDLYTPMG--QCCLHTGFMG 859
DB 932 CL-----CTKGYEGRDCAIMTDDC--ASFPCCQNGGTCLD---GIGDYSCLCVGDFG 978
OY 860 TACEL-----CMHGRF-----GPDCCPRGSEHGQCD 886
DB 979 KHCIEDINECLSQPCQNGATCSQYVNSYCTGCPLEFGSINGTCOTNDEDTRESSCLNGSC 1038
OY 887 EGTGSGEFLCTGTMFAASCDPTPAFAVC-TPACSVHATCTENN--TCVCNLINEGCG 942
DB 1039 DGIINGY-NCSCLAGSYGANCQYK---LNKCDNSPLCLNGATCHRONNRYTHCHPSGTTGR- 1093
OY 943 ITCT-VVDFCKQNNNGCAKAVKCSOKGTVGSSCKKGYKGDYSC-TEIDPCADGV---- 996
DB 1094 -CSEXYVDMCGGS--PCENGAATCSQMKHOFSCCKSGAGMTGK--LCDVQYITSCODADBRKG 1148
OY 997 -----NGGCHENATCMTGRGKHKCECKSHIYGVGDYDCEPEOLPLDRCLQDNQCHP 1048
DB 1149 LSLRLQNNNGTCKDYG-----NSHVYCQSQGYAGS--YCKQK--IDEC--QSQPCQN 1194
OY 1049 DASCADLYFDQITVGVFHLRSLPGQYKLFDPKAKEACAKEAATATYNOLSYAQRKAYHL 1108
DB 1195 GGTGRDL-----IGAVEDQCRGQVFGQNCCLNIDCAINP----- 1229
OY 1109 CSAGWLESGRVAVPTTYASQKGCANVGTVDYGSRANKSEMDVFCY-----RMKDVN 1161
DB 1230 CQNGGTCHDRV-----MNFSCSPGPTGMII--CEINKDKCRKAGACHNNNGSCTIDRVGFE 1282
OY 1162 CTCKAGYVGDGFSGSGNLLQVL 1183
DB 1283 CYCQPGFV--GARCEGDINECL 1302

```

RESULT 7

ID	FBNI_PIG	STANDARD	PRT	2871 AA.
AC	09TV36;			
DT	16-OCT-2001 (Rel. 40; Created)			
DT	16-OCT-2001 (Rel. 40; Last sequence update)			
DT	16-OCT-2001 (Rel. 40; Last annotation update)			
DE	Fibrillin 1 precursor.			
GN	FBNI.			
OS	Sus scrofa (pig).			
OC	Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.			
OX	NCBI_TaxID=9823;			
RN	(1)			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Lung;			
RX	MEDLINE=99156858; PubMed=10036187;			
RA	Bieri N.J., Eldadah Z.A., Moore C.S., Stetten G., Spencer F.,			
RA	Dietz H.C.;			
RT	Revised genomic organization of FBNI and significance for regulated			
RT	gene expression."			
RL	Genomics 56:70-77(1999).			
CC	- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS			
CC	THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS PROVIDE			
CC	LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.			
CC	- PFM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER			
CC	FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE			
CC	MICROFIBRILS.			
CC	- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING			
CC	EGF-LIKE DOMAINS.			
CC	- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.			
CC	CC This swiss-prot entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation - its			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			

CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch)			
CC	-----			
DR	EMBL: AF073800; AAD50328.1; -			
DR	HSSP: P35555; IART			
DR	InterPro: IPR000152; Asx_hydroxyl.			
DR	InterPro: IPR000561; EGF-like.			
DR	InterPro: IPR001881; EGF_Ca.			
DR	InterPro: IPR001438; EGF_II.			
DR	InterPro: IPR002212; TB.			
DR	Pfam: PF00008; EGF_45.			
DR	Pfam: PF00683; TB; 9.			
DR	PRINTS: PR00010; EGFBL00D.			
DR	SMART: SM00179; EGF_CA; 40.			
DR	SMART: SM00001; EGF_like; 6.			
DR	PROSITE: PS00010; ASX_HYDROXYL; 41.			
DR	PROSITE: PS00022; EGF_1; UNKNOWN_2.			
DR	PROSITE: PS01187; EGF_CA; 41.			
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;			
KW	Repeat; Signal; Multigene family.			
FT	SIGNAL	1	27	POTENTIAL. 1.
FT	CHAIN	28	2871	
FT	DOMAIN	81	112	EGF-LIKE 1, NON-CALCIUM BINDING.
FT	DOMAIN	115	146	EGF-LIKE 2, NON-CALCIUM BINDING.
FT	DOMAIN	147	178	EGF-LIKE 3, NON-CALCIUM BINDING.
FT	DOMAIN	246	287	EGF-LIKE 4, CALCIUM-BINDING.
FT	DOMAIN	288	329	EGF-LIKE 5, CALCIUM-BINDING.
FT	REPEAT	330	390	TEFBP 1.
FT	DOMAIN	392	446	PRO-RICH.
FT	DOMAIN	449	489	EGF-LIKE 6, NON-CALCIUM BINDING.
FT	DOMAIN	490	529	EGF-LIKE 7, CALCIUM-BINDING.
FT	DOMAIN	530	571	EGF-LIKE 8, CALCIUM-BINDING.
FT	DOMAIN	572	612	EGF-LIKE 9, CALCIUM-BINDING.
FT	DOMAIN	613	653	EGF-LIKE 10, CALCIUM-BINDING.
FT	REPEAT	654	722	TEFBP 2.
FT	DOMAIN	723	764	EGF-LIKE 11, CALCIUM-BINDING.
FT	DOMAIN	765	806	EGF-LIKE 12, CALCIUM-BINDING.
FT	DOMAIN	807	846	EGF-LIKE 13, CALCIUM-BINDING.
FT	DOMAIN	910	951	EGF-LIKE 14, CALCIUM-BINDING.
FT	REPEAT	952	1027	TEFBP 3.
FT	DOMAIN	1028	1069	EGF-LIKE 15, CALCIUM-BINDING.
FT	DOMAIN	1070	1112	EGF-LIKE 16, CALCIUM-BINDING.
FT	DOMAIN	1113	1154	EGF-LIKE 17, CALCIUM-BINDING.
FT	DOMAIN	1155	1196	EGF-LIKE 18, CALCIUM-BINDING.
FT	DOMAIN	1197	1237	EGF-LIKE 19, CALCIUM-BINDING.
FT	DOMAIN	1238	1279	EGF-LIKE 20, CALCIUM-BINDING.
FT	DOMAIN	1280	1321	EGF-LIKE 21, CALCIUM-BINDING.
FT	DOMAIN	1322	1362	EGF-LIKE 22, CALCIUM-BINDING.
FT	DOMAIN	1363	1403	EGF-LIKE 23, CALCIUM-BINDING.
FT	DOMAIN	1404	1445	EGF-LIKE 24, CALCIUM-BINDING.
FT	DOMAIN	1446	1486	EGF-LIKE 25, CALCIUM-BINDING.
FT	DOMAIN	1487	1527	EGF-LIKE 26, CALCIUM-BINDING.
FT	REPEAT	1528	1605	TEFBP 4.
FT	DOMAIN	1606	1647	EGF-LIKE 27, CALCIUM-BINDING.
FT	DOMAIN	1648	1688	EGF-LIKE 28, CALCIUM-BINDING.
FT	REPEAT	1689	1765	TEFBP 5.
FT	DOMAIN	1766	1807	EGF-LIKE 29, CALCIUM-BINDING.
FT	DOMAIN	1808	1848	EGF-LIKE 30, CALCIUM-BINDING.
FT	DOMAIN	1849	1890	EGF-LIKE 31, CALCIUM-BINDING.
FT	DOMAIN	1891	1929	EGF-LIKE 32, CALCIUM-BINDING.
FT	DOMAIN	1930	1972	EGF-LIKE 33, CALCIUM-BINDING.
FT	DOMAIN	1973	2012	EGF-LIKE 34, CALCIUM-BINDING.
FT	DOMAIN	2013	2054	EGF-LIKE 35, CALCIUM-BINDING.
FT	REPEAT	2055	2126	TEFBP 6.
FT	DOMAIN	2127	2165	EGF-LIKE 36, CALCIUM-BINDING.
FT	DOMAIN	2166	2205	EGF-LIKE 37, CALCIUM-BINDING.
FT	DOMAIN	2206	2246	EGF-LIKE 38, CALCIUM-BINDING.
FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING.
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCIUM-BINDING.
FT	REPEAT	2333	2401	TEFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING.

FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.
FT	DISULFID	1326	1339	BY SIMILARITY.
FT	DISULFID	1333	1348	BY SIMILARITY.
FT	DISULFID	1350	1361	BY SIMILARITY.
FT	DISULFID	1367	1380	BY SIMILARITY.
FT	DISULFID	1374	1389	BY SIMILARITY.
FT	DISULFID	1391	1402	BY SIMILARITY.

Query Match	6.2%	Score 483.5;	DB 1;	Length 2871;
Best Local Similarity	22.3%	Pred. No. 7.9e-26;		
Matches 259;	Conservative 116;	Mismatches 342;	Indels 443;	Gaps 69

1012 PGFATKEITNGKPF--EKDINECKMIPNLCTHGKCRNTI--GSFKCRDSDGFALDS---E 106

Db 1421 --LY---APGCGHCEYDMG 1434

CC modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce>)

CC or send an email to license@isb.sib.ch).

DR EMBL; M73980; AAA60614.1; -
DR HSSP; P00740; IIXA.
DR MIM; 190198; -
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-Like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00003; ank; 6.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00001; EGF_Like; 13.
DR SMART; SM00004; NL; 2.
DR PROSITE; PSS0086; ANK_REPEAT; 4.
DR PROSITE; PSS0297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 20.
DR PROSITE; PS00022; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 26.
DR PROSITE; PS01187; EGF_CA; 18.
KW Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 18
FT CHAIN 19 >2444
FT DOMAIN 19 1736
FT TRANSLEM 1737 1757
FT DOMAIN 1738 >2444
FT DOMAIN 20 58
FT DOMAIN 59 99
FT DOMAIN 102 139
FT DOMAIN 140 176
FT DOMAIN 178 216
FT DOMAIN 218 235
FT DOMAIN 257 293
FT DOMAIN 295 333
FT DOMAIN 335 371
FT DOMAIN 372 410
FT DOMAIN 412 450
FT DOMAIN 452 488
FT DOMAIN 490 526
FT DOMAIN 528 564
FT DOMAIN 566 601
FT DOMAIN 603 639
FT DOMAIN 641 676
FT DOMAIN 678 714
FT DOMAIN 716 751
FT DOMAIN 753 789
FT DOMAIN 791 827
FT DOMAIN 829 868
FT DOMAIN 870 906
FT DOMAIN 908 944
FT DOMAIN 946 982
FT DOMAIN 984 1020
FT DOMAIN 1022 1058
FT DOMAIN 1060 1096
FT DOMAIN 1098 1144
FT DOMAIN 1146 1182
FT DOMAIN 1184 1220
FT DOMAIN 1222 1266
FT DOMAIN 1268 1306
FT DOMAIN 1308 1347
FT DOMAIN 1349 1385
FT DOMAIN 1388 1427
FT DOMAIN 1446 1481
FT REPEAT 1482 1523
FT REPEAT 1524 1563
FT REPEAT 1568 1597
FT REPEAT 1591 1991
FT REPEAT 1995 2024

FT REPEAT 2028 2057 ANK 4.
FT REPEAT 2061 2090 ANK 5.
FT DOMAIN 1576 1579 POLY-VAL.
FT DOMAIN 1662 1665 POLY-ARG.
FT DOMAIN 1729 1732 POLY-PRO.
FT DOMAIN 1741 1744 POLY-ALA.
FT DOMAIN 1902 1905 POLY-GLU.
FT DOMAIN 2260 2263 POLY-GLY.
FT DOMAIN 2404 2407 POLY-GLN.
FT DOMAIN 2411 2418 POLY-PRO.
FT DISULFID 24 37 BY SIMILARITY.
FT DISULFID 31 46 BY SIMILARITY.
FT DISULFID 48 57 BY SIMILARITY.
FT DISULFID 63 74 BY SIMILARITY.
FT DISULFID 68 87 BY SIMILARITY.
FT DISULFID 89 98 BY SIMILARITY.
FT DISULFID 106 117 BY SIMILARITY.
FT DISULFID 111 127 BY SIMILARITY.
FT DISULFID 129 138 BY SIMILARITY.
FT DISULFID 144 155 BY SIMILARITY.
FT DISULFID 149 164 BY SIMILARITY.
FT DISULFID 166 175 BY SIMILARITY.
FT DISULFID 182 195 BY SIMILARITY.
FT DISULFID 189 204 BY SIMILARITY.
FT DISULFID 206 215 BY SIMILARITY.
FT DISULFID 222 233 BY SIMILARITY.
FT DISULFID 227 243 BY SIMILARITY.
FT DISULFID 245 254 BY SIMILARITY.
FT DISULFID 261 272 BY SIMILARITY.
FT DISULFID 266 281 BY SIMILARITY.
FT DISULFID 283 292 BY SIMILARITY.
FT DISULFID 299 312 BY SIMILARITY.
FT DISULFID 306 321 BY SIMILARITY.
FT DISULFID 323 332 BY SIMILARITY.
FT DISULFID 339 350 BY SIMILARITY.
FT DISULFID 344 359 BY SIMILARITY.
FT DISULFID 361 370 BY SIMILARITY.
FT DISULFID 376 387 BY SIMILARITY.
FT DISULFID 381 398 BY SIMILARITY.
FT DISULFID 400 409 BY SIMILARITY.
FT DISULFID 416 429 BY SIMILARITY.
FT DISULFID 423 438 BY SIMILARITY.
FT DISULFID 440 449 BY SIMILARITY.
FT DISULFID 456 467 BY SIMILARITY.
FT DISULFID 461 476 BY SIMILARITY.
FT DISULFID 478 487 BY SIMILARITY.
FT DISULFID 494 505 BY SIMILARITY.
FT DISULFID 499 514 BY SIMILARITY.
FT DISULFID 516 525 BY SIMILARITY.
FT DISULFID 532 543 BY SIMILARITY.
FT DISULFID 537 552 BY SIMILARITY.
FT DISULFID 554 563 BY SIMILARITY.
FT DISULFID 570 580 BY SIMILARITY.
FT DISULFID 575 589 BY SIMILARITY.
FT DISULFID 591 600 BY SIMILARITY.
FT DISULFID 607 618 BY SIMILARITY.
FT DISULFID 612 627 BY SIMILARITY.
FT DISULFID 629 638 BY SIMILARITY.
FT DISULFID 645 655 BY SIMILARITY.
FT DISULFID 650 664 BY SIMILARITY.
FT DISULFID 666 675 BY SIMILARITY.
FT DISULFID 682 693 BY SIMILARITY.
FT DISULFID 687 702 BY SIMILARITY.
FT DISULFID 704 713 BY SIMILARITY.
FT DISULFID 720 730 BY SIMILARITY.
FT DISULFID 725 739 BY SIMILARITY.
FT DISULFID 741 750 BY SIMILARITY.
FT DISULFID 757 768 BY SIMILARITY.
FT DISULFID 762 777 BY SIMILARITY.
FT DISULFID 779 788 BY SIMILARITY.
FT DISULFID 795 806 BY SIMILARITY.
FT DISULFID 800 815 BY SIMILARITY.
FT DISULFID 817 826 BY SIMILARITY.

Query Match 6.1% Score 482.5; DB 1; Length 2444;
 Best Local Similarity 21.7%; Pred. No. 7.5e-26;
 Matches 269; Conservative 120; Mismatches 445; Indels 407; Gaps 75;

127 EKVLEIKNKNNDNNTIVIGEC--GKCSQAPRPLE-----TKPLRETR----- 169
 256 ENIDDCGNCKNG-----GACVDVNTYNCPCPEMTGQYCTEDVDDECOLMNAQNG 309
 170 -----KCIYSIFMGKRSVFIGCQPCVRTI---ITRACMLASLHNAK----- 210
 310 GTCHNTHGIVNCVNCWMTGE-----DCSENIDDCASACFHCATCHDRASRYCE 360
 211 -PABGEVKMCALGTASY-----WDGVNGTGTCCGLGFNGTAC---ETCTEGK 254
 361 CPHGRTGLLCLNDACISNPNCSNCDTNFVNGKALICTCPSTGYTPACSDQVDECSLG 419
 255 YGHCADACSCVHRCGSGPLGDSGCCDVGMRGVKCDMETTTNCG-NGTCHTSANCLLD 313
 420 -ANPCEHAGKCN-----TLGSFECCLGTYGTPRC--EIDVNECVSNPCQNDATC-LD 469
 314 PDGKASCKCAAGFRNGTVCATAINA---CETS---NGCSTKADCKRTTPGNRYCVCKAG 367
 470 QIGFQCMCPGYRG---VHCEVNTDECASSPCLHNGRCLDKIN-----EPCCPCPTG 519
 368 YTGDIYC-LEINPCLNHHGCGDRNAECTQTGPNAVANCNCLPKYTGDKYCSLINVCLTN 426
 520 FTGH--LCQYDVDECAST--PCKNGAKCLD-GPNTYTCVCTEGYTG-----THCEVD 566
 427 NGGSPAPFCNYTQDRI---CTCKPDYTGDIIVRGSIYGLPKRNPSTSYFFPQOE 482
 567 IDECDPPD-CHYGSCKDGVATFTCLCRPGYTGHH--CETNI-NECSSQPCRLRGTCDDPD 622
 483 HA-----VRELAGRPPTVFAPLSSSFNHEPRIKDWDQGLMSQVLRYHYVGCQQLLDN 537
 623 NAYLCFLCKGTGP-----NCEINLD-----CASSPCDS 652
 538 LKVTATSTLQGEFVSISVSODTIVINNEAKVLSDDITSTNGVINVHDKLSPPNLLTP 597
 653 ---GTCLDKIDGYECACEPEGYTGSNCNSNIDECAGNPNCHNG-----TC 693
 598 KDALGRVLONTTVAANHGYTKFKSKLLJODSGLSVITDSIHTPYTVWPTDKALEALPPE 657
 694 EDGI-----NGFTCRCEGY-----HDPYCLSEVNECNSNCPV----- 726
 658 QODFLFMDNKKDKLSYLFKHV-----IRDSKALSDLPKRSASWKTLOGSELSVR 707
 727 -----HGACRDSLNGY-KCDCCDPGWSGTNDINNNECESPCVNGGCKMT--SGIVCT 778
 708 CGTG-----SDIGELFLNEQMCRFIHRGLFDVGAAYGIDCLLIMPTLGGRCDTFTTF 760
 779 CREFGSPNCQNTINIECASNP--C--LNKGTICIDVAGYKNCNCLL--PYTGATCEVVLAP 832
 761 DIPREC---GSC-----IFTPKCPILSKRKGVK-----KKCIYNPLPFRNNEGC 802
 833 CAPSPCHNGCROSEDEYSEFSVCP---TAGAKGOTCEVDINECVLSPC--RHGASC 885

QY 803 QNLCTVVIQTPRCHGYEMPDQACPGPD-----TPCNRNGKCRD-LYTPMGOC 851
 DB 886 QN-----THGXRYCHQAGYSGRNCEITDLDORPRPCHNGSGCTGTGINTAF--C 932
 QY 852 LCHTFNGTAGE-----LCMHGRFG-----PDCPRS 878
 DB 933 DCLGFGRTFCEEDINECASDPCRNKNGANTDCVDSYTCYCPAGSGIHCENNFTDCTESS 992
 QY 879 CSEHGQCEDEGTGSGECLCTEFGMTAASCDPTTAVPAAC--TPACSVHATCTEN---NVCV 934
 DB 993 CFNGGTCVDGI-NSFTCLCPRGFTGYSYQ---HVNNECDSPCLLGGTCODGRGLHRCCTC 1048
 QY 935 NLNNEGDCITC-TVWDPCCKNNGGCAKAVKCSQGTQVSCSCSKRGYK-----DGYSCIE 988
 DB 1049 PQGTGP--NQNVLVHMC--DSSPCKNKGCKWOTHTQYRCBCPSGWTGLYCDVSVSC-E 1103
 QY 989 IDPCADGVNCG--CHEHATCMTGPKNHCKECSKSHYVG---DQVD-CEPQQLPLDRCLQ 1041
 DB 1104 VAAOROGVDVARLCOHGLICVDAG-NTHHRCQAGYTGSGYCEDLVDECSF----- 1152
 QY 1042 DNGCHPDASCADLYFDQDTYGVFHLKSLPQGYKLTFDKAEACAKKATATATYNQLSYA 1101
 DB 1153 --SPQNGATCTDY-----LGGYSCKCVAGYHGVNCSSEIDECLS----- 1190
 QY 1102 QKARYHLCSAGMLESGRVAYPTTASQKCGANVGI-----VDYGRANKSEM 1149
 DB 1191 -----HPCQNG---GTCLDLPLNTYK--SCPRGTQGVHCEINVDONCPVDPVSNPK--- 1238
 QY 1150 WDVFCY-----RMKDVNCTCKAGYVGDGFSGSLLOVL 1183
 DB 1239 ---CFNNGTCVDPQGVGYSCTCPPGFGE--RCGADVNECL 1273

RESULT 9
 FBNI_BOVIN
 ID FBNI_BOVIN STANDARD; PRT; 2871 AA.
 AC P98133;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Fibrillin 1 precursor (MP340).
 GN FBNI.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_Taxid=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=SKIN;
 RX MEDLINE=95137597; PubMed=7835900;
 RA Tilstra D.J., Potter K.A., Byers P.H.;
 RT "Sequence of the coding region of the bovine fibrillin cDNA and
 RT localization to bovine chromosome 10.";
 RL Genomics 23:480-485(1994).
 RN [2]
 RP PARTIAL SEQUENCE.
 RX MEDLINE=96132851; PubMed=8557636;
 RA Gibson M.A., Hatzinikolas G., Kumaratilake J.S., Sandberg L.B.,
 RA Nicholl J.K., Sutherland G.R., Cleary E.G.;
 RT "Further characterization of proteins associated with elastic fiber
 RT microfibrils including the molecular cloning of MAGP-2 (MP25).";
 RL J Biol. Chem. 271:1096-1103(1996).
 CC -!- FUNCTION: STRUCTURAL COMPONENT OF CONNECTIVE TISSUE MICROFIBRILS
 CC THAT BINDS CALCIUM. FIBRILLIN-1-CONTAINING MICROFIBRILS
 CC LONG-TERM FORCE BEARING STRUCTURAL SUPPORT.
 CC -!- PTM: FORMS INTERMOLECULAR DISULFIDE BONDS EITHER WITH OTHER
 CC FIBRILLIN-1 MOLECULES OR WITH OTHER COMPONENTS OF THE
 CC MICROFIBRILS.
 CC -!- SIMILARITY: CONTAINS 43 CALCIUM-BINDING AND 4 NON-CALCIUM BINDING
 CC EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 7 TGF-BETA BINDING PROTEIN DOMAINS.

CC		CC This Swiss-Prot entry is copyright. It is produced through a collaboration -
CC		CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC		CC the European Bioinformatics Institute. There are no restrictions on its
CC		CC use by non-profit institutions as long as its content is in no way
CC		CC modified and this statement is not removed. Usage by and for commercial
CC		CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC		CC or send an email to license@isb-sib.ch)
CC		-----
DR	EMBL; L28748; AAA74122.1; ..	
DR	HSP; P35555; IAPJ.	
DR	InterPro: IPRO00152; Asx_hydroxyl.	
DR	InterPro: IPRO00561; EGF-like.	
DR	InterPro: IPRO01881; EGF_Ca.	
DR	InterPro: IPRO01438; EGF_IT.	
DR	InterPro: IPRO02212; TB.	
DR	Pfam; PF00683; EGF; 46.	
DR	Pfam; PF00683; TB; 9.	
DR	PRINTS; PRO0010; EGFBLD.	
DR	SMART; SM00179; EGF_CA; 42.	
DR	SMART; SM00001; EGF_like; 4.	
DR	PROSITE; PS00010; ASX_HYDROXYL; 43.	
DR	PROSITE; PS00022; EGF_1; 2.	
DR	PROSITE; PS01186; EGF_2; 38.	
DR	PROSITE; PS01187; EGF_CA; 45.	
KW	Extracellular matrix; Calcium-binding; Glycoprotein; EGF-like domain;	
KW	Repeat; Signal; Multigene family.	
KW	SIGNAL	1 27
FT	CHAIN	28 2871
FT	DOMAIN	81 112
FT	DOMAIN	115 146
FT	DOMAIN	147 178
FT	DOMAIN	246 287
FT	DOMAIN	288 329
FT	DOMAIN	330 390
FT	REPEAT	1.
FT	DOMAIN	392 446
FT	DOMAIN	449 489
FT	DOMAIN	490 529
FT	DOMAIN	530 571
FT	DOMAIN	572 612
FT	DOMAIN	613 653
FT	DOMAIN	654 722
FT	REPEAT	723 764
FT	DOMAIN	765 806
FT	DOMAIN	807 846
FT	DOMAIN	910 951
FT	DOMAIN	952 1027
FT	REPEAT	1028 1069
FT	DOMAIN	1070 1112
FT	DOMAIN	1113 1154
FT	DOMAIN	1155 1196
FT	DOMAIN	1197 1237
FT	DOMAIN	1238 1279
FT	DOMAIN	1280 1321
FT	DOMAIN	1322 1362
FT	DOMAIN	1363 1403
FT	DOMAIN	1404 1445
FT	DOMAIN	1446 1486
FT	DOMAIN	1487 1527
FT	REPEAT	1528 1605
FT	DOMAIN	1606 1647
FT	DOMAIN	1648 1688
FT	REPEAT	1689 1765
FT	DOMAIN	1766 1807
FT	DOMAIN	1808 1848
FT	DOMAIN	1849 1890
FT	DOMAIN	1891 1929
FT	DOMAIN	1930 1972
FT	DOMAIN	1973 2012
FT	DOMAIN	2013 2054
FT	REPEAT	2055 2126
FT	DOMAIN	2127 2165
FT	DOMAIN	2166 2205
FT	DOMAIN	2206 2246

FT	DOMAIN	2247	2290	EGF-LIKE 39, CALCIUM-BINDING
FT	DOMAIN	2291	2332	EGF-LIKE 40, CALCIUM-BINDING
FT	REPEAT	2333	2401	TGFBP 7.
FT	DOMAIN	2402	2443	EGF-LIKE 41, CALCIUM-BINDING
FT	DOMAIN	2444	2484	EGF-LIKE 42, CALCIUM-BINDING
FT	DOMAIN	2485	2523	EGF-LIKE 43, CALCIUM-BINDING
FT	DOMAIN	2524	2566	EGF-LIKE 44, CALCIUM-BINDING
FT	DOMAIN	2567	2606	EGF-LIKE 45, CALCIUM-BINDING
FT	DOMAIN	2607	2647	EGF-LIKE 46, CALCIUM-BINDING
FT	DOMAIN	2648	2687	EGF-LIKE 47, CALCIUM-BINDING
FT	DISULFID	85	94	BY SIMILARITY.
FT	DISULFID	89	100	BY SIMILARITY.
FT	DISULFID	102	111	BY SIMILARITY.
FT	DISULFID	119	129	BY SIMILARITY.
FT	DISULFID	123	134	BY SIMILARITY.
FT	DISULFID	136	145	BY SIMILARITY.
FT	DISULFID	150	160	BY SIMILARITY.
FT	DISULFID	154	166	BY SIMILARITY.
FT	DISULFID	168	177	BY SIMILARITY.
FT	DISULFID	250	262	BY SIMILARITY.
FT	DISULFID	257	271	BY SIMILARITY.
FT	DISULFID	273	286	BY SIMILARITY.
FT	DISULFID	292	304	BY SIMILARITY.
FT	DISULFID	299	313	BY SIMILARITY.
FT	DISULFID	315	328	BY SIMILARITY.
FT	DISULFID	453	465	BY SIMILARITY.
FT	DISULFID	460	474	BY SIMILARITY.
FT	DISULFID	476	488	BY SIMILARITY.
FT	DISULFID	494	504	BY SIMILARITY.
FT	DISULFID	499	513	BY SIMILARITY.
FT	DISULFID	515	528	BY SIMILARITY.
FT	DISULFID	534	546	BY SIMILARITY.
FT	DISULFID	541	555	BY SIMILARITY.
FT	DISULFID	557	570	BY SIMILARITY.
FT	DISULFID	576	587	BY SIMILARITY.
FT	DISULFID	582	596	BY SIMILARITY.
FT	DISULFID	598	611	BY SIMILARITY.
FT	DISULFID	617	628	BY SIMILARITY.
FT	DISULFID	623	637	BY SIMILARITY.
FT	DISULFID	639	652	BY SIMILARITY.
FT	DISULFID	727	739	BY SIMILARITY.
FT	DISULFID	734	748	BY SIMILARITY.
FT	DISULFID	750	763	BY SIMILARITY.
FT	DISULFID	769	781	BY SIMILARITY.
FT	DISULFID	776	790	BY SIMILARITY.
FT	DISULFID	792	805	BY SIMILARITY.
FT	DISULFID	811	821	BY SIMILARITY.
FT	DISULFID	816	830	BY SIMILARITY.
FT	DISULFID	832	845	BY SIMILARITY.
FT	DISULFID	914	926	BY SIMILARITY.
FT	DISULFID	921	935	BY SIMILARITY.
FT	DISULFID	937	950	BY SIMILARITY.
FT	DISULFID	1032	1044	BY SIMILARITY.
FT	DISULFID	1039	1053	BY SIMILARITY.
FT	DISULFID	1055	1068	BY SIMILARITY.
FT	DISULFID	1074	1086	BY SIMILARITY.
FT	DISULFID	1081	1095	BY SIMILARITY.
FT	DISULFID	1097	1111	BY SIMILARITY.
FT	DISULFID	1117	1129	BY SIMILARITY.
FT	DISULFID	1124	1138	BY SIMILARITY.
FT	DISULFID	1140	1153	BY SIMILARITY.
FT	DISULFID	1159	1171	BY SIMILARITY.
FT	DISULFID	1166	1180	BY SIMILARITY.
FT	DISULFID	1182	1195	BY SIMILARITY.
FT	DISULFID	1201	1212	BY SIMILARITY.
FT	DISULFID	1208	1221	BY SIMILARITY.
FT	DISULFID	1223	1236	BY SIMILARITY.
FT	DISULFID	1242	1254	BY SIMILARITY.
FT	DISULFID	1249	1263	BY SIMILARITY.
FT	DISULFID	1265	1278	BY SIMILARITY.
FT	DISULFID	1284	1296	BY SIMILARITY.
FT	DISULFID	1291	1305	BY SIMILARITY.
FT	DISULFID	1307	1320	BY SIMILARITY.

Query Match Best Local Similarity 20.8%; Pred. No. 1,2e-25; Matches 276; Conservative 128; Mismatches 383; Indels 542; Gaps 77;

125 GLEKLELORRCD---NNDTIIVRG-----ECGKCSQOAPCPLETPPLRET 168
1100 GYESGFMMKMKMDIDECQRPDLCRGGVCLNTEGSYRCECPGQOLAP---NISACIDI 1156
169 KCIYSIYFMGRSRYFIC-OPOCVRTIITRACMLASLAHNAKPAPEGVKKALCTASVM 227
1157 NECELSAIL-----CPHGRCVNLIGKYOACNPGYHS---TPDRLFCVDIDECSTIM 1204
228 DG-----VNCTGT--COGGLGF-----NGTACE--TCTE--GKGYIH 258
1205 NGGCEFTCTNSEGSEYECSCOPGFALMPDQSCITDIDECEDNPNICDGGQCTNIPGEYRCL 1264
259 CDQA-----CS-----CVHGRCSOGPLGDSGDDQVGRGVK----C--- 291
1265 CYDGFMASDEKKTCDVNECOLNPNICLSGTC-ENTKGSFICHOMGSGKKGKTCCTDI 1323
292 -DMEITTDNC--NGTCHTSANCLDDPDGKASCKCAAGFRGNVTGTAINAGETSNCGCST 348
1324 NECEIGAHNCORHVAVCTMTA-----GSFKGSCSPGWICIGIKTDLDECSNGTHMSQ 1376
349 KADCKRTTPGRKVCYCKAGTYGDIIVCLEINPCLENHGGCDRMAECTGTGPNOAVCNC-- 406
1377 HADCKNTM-GSYRCLCKEGYTGDEFTCTDLDECSENLMICG-NGQCL-NAPGVGRCEGDM 1433
407 --LPKYGDGVK---CSLINVCL-----TNGGCSPPAFNC 436
1434 GFVP--SADGRACEDIDECSLPNICVPGTCHNLPGLFRCEGEIYELDRSGNCTDVNEC 1491
437 -----NYTEDODRITCKPDY-----TG 454
1492 LDPTTCISGNCVNTPGSYT-----CDCPDPFELNPTRVGCVDRHSAGCYLDIRRGDNG 1545
455 D-----GIYCR-GSIYV--ELPKNPSTSOYFPQLOEHAHVRELAFG--- 492
1546 DTACSNEIGVSVKASCCSLGKAMGTPCELCPPVNTSEY-----KILCGGEGF 1595
493 ---PFTVAFPLSSSFNHPRIKMDQOGLMSOVLRYHVVGQOL--LLDNLKVTTSATTL 547
1596 RPNPITVY-----LEDIDE-----COELPGLCOGGGCLNTEGTF 1629
548 OGE-PVSTSVSODTVF-----INNAKVLUSDIIISNGVIHVID 585
1630 QCRCPGYLLMEDIRVYCDVNECEPPTGICGPTCVNTYAGNTTICPPRYMOYVNGSNMCD 1689
586 --KLSPKNLLITPRDALGRVQLONTTVAAHNGYTKFSKLLIODSGLSVTDSHTPTVY 643
1690 MRRSLCYRNAYADNQTCGELLFNTKMKCCSYN-----IGR 1727
644 FWPIDKALEALPPEGODPLFQDNKMDKLKYLKPHVIRDSKALASDLPRASAKMTLOSSE 703
1728 AM--NKPCGQCP-----IPSTDEPVLGCSQ 1751

704 -----LSVRCTGSDIGELFLNEQMKRFTIHRQLLFDVGAVAGIDCLMNPVLGRCDPTT 758
1752 RPFGEIDYLTGLPVIDE-----CR----- 1771
759 TFDIDGEC--GSCI-----PTPKPLKSPKGVKKKCIYN-PLPRRNVGCGQN--LCTV 808
1772 --EIPGVCENGVCIMMVSFRCCEPV-----GFYFNKDLIACVEDIDECQNPVCO- 1819
809 VIQPRGCH--GYFPPDOACGCPDT--PCNNRGMCRDLATTPMGOCCLHTGFSTACE 863
1820 --RNAECINTGASRYCDCK--PGYRFTSTGQCDNNECOET----- 1856
864 LCMHGRFEPDQCPRSCSEHGGQCEGITSGBELCTEFTGTAAACDPRTAVFAVCPACSVH 923
1857 -----PNICS-HGQCIDIV--GSFYCLCHTGRKTANDQYMCIDLINCEBDAQCN 1902
924 ATCTENNT-----CYCN-----LNYEDGITTCTVDFPCKONNGCGAKVAKCSOKGTQVSCS 974
1903 GTC--RNTIGSFNCRGNHGFILSHNND--CIDVDECATGNMNLGRNQCINTVGSFQCQ 1957
975 CKKGK--GDGYSCTEI-----DP--CADGVNCGCHENATCRMTGPRGHKCRSHYVD 1025
1958 CMEGYEVAPDGRCTVDINECLDPRKCAPG-----TCQNL-DGSTKCLCPPGYSLQ 2007
1026 GVDCPEQLPLDRCLQDNQCHPDASCADLYFQDITVGVFHLRSPDLGOYKLTDFKAKBAC 1085
2008 NDKCE--DIDCEYERPEIC-ALQTCN-----TEBSFKCLCPDQ--FSLSTGKRC 2053
1086 AKEAATYATYNOLSYAKAKYHLCSAGMLESGRAYPTT--YASOKGANYVG----- 1136
2054 -----DULRMSYCYAKE-EGGKCSPPKSRHNSKOECCALKGBGWDPC 2096
1137 -----IVDYS-----RANKEMMDVPCYRNMKDVN-----CTCKA 1166
2097 ELCPPEPDEAPRQIDPYSGIITVGPDSAVMDMBCKBPV--CKHGQCIINTGSIYRCECP 2155
1167 GYVGDGFSQ 1175
2156 GYIIQGNEC 2164

RESULT 10
NOTC_BRARE STANDARD: PRT: 2437 AA.
AC P46530:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neurogenic locus notch homolog protein precursor.
GN NOTCH1 OR NOTCH1A.
OS Brachydanio rerio (zebrafish) (zebra danio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Ostariophysi;
OC Cypriniformes; Cyprinidae; Danio.
OX NCBI_TaxID=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=94128602; PubMed=8297791;
RA Bierkamp C., Campos-Ortega J.A.;
RT "A zebrafish homologue of the Drosophila neurogenic gene Notch and
RT its pattern of transcription during early embryogenesis.";
RL Mech. Dev. 43:87-100(1993)
CC -!- FUNCTION: IMPLICATED IN CELL FATE SPECIFICATIONS DURING
CC EMBRYO DEVELOPMENT. MAY BE INVOLVED IN THE FORMATION OF THE
CC NEURAL PLATE, NOTOCHORD AND BRAIN VESICLES.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN ALL CELLS IN PREGASTRULATION
CC STAGES. DURING GASTRULATION IS DIFFERENTIALLY EXPRESSED
CC ACCUMULATING PREDOMINANTLY IN THE PRECHORDAL MESODERM AND
CC NOTOCHORD. AT THE END OF GASTRULATION, EXPRESSED ALONG THE
CC ANTERIOR-POSTERIOR AXIS INCLUDING THE DEVELOPING NEURAL PLATE
CC AND DIFFERENTIATING MESODERM. ALSO PRESENT IN THE DEVELOPING

FT DISULFID 718 728 BY SIMILARITY.
FT DISULFID 723 737 BY SIMILARITY.
FT DISULFID 739 748 BY SIMILARITY.
FT DISULFID 755 766 BY SIMILARITY.
FT DISULFID 760 775 BY SIMILARITY.
FT DISULFID 777 786 BY SIMILARITY.
FT DISULFID 793 804 BY SIMILARITY.
FT DISULFID 798 813 BY SIMILARITY.
FT DISULFID 815 824 BY SIMILARITY.
FT DISULFID 831 842 BY SIMILARITY.
FT DISULFID 836 853 BY SIMILARITY.
FT DISULFID 855 864 BY SIMILARITY.
FT DISULFID 871 882 BY SIMILARITY.
FT DISULFID 876 891 BY SIMILARITY.
FT DISULFID 893 902 BY SIMILARITY.
FT DISULFID 909 920 BY SIMILARITY.
FT DISULFID 914 929 BY SIMILARITY.
FT DISULFID 931 940 BY SIMILARITY.
FT DISULFID 947 958 BY SIMILARITY.
FT DISULFID 952 967 BY SIMILARITY.
FT DISULFID 969 978 BY SIMILARITY.
FT DISULFID 1023 1034 BY SIMILARITY.
FT DISULFID 1028 1043 BY SIMILARITY.
FT DISULFID 1045 1054 BY SIMILARITY.

Query Match 6.0%; Score 473.5; DB 1; Length 2437;
Best local Similarity 20.3%; Pred. No. 3,3e-25;

Matches 303; Conservative 123; Mismatches 447; Indels 617; Gaps 84;

OY 135 NRCNDNDIIVKEGC-----KCSQQA-PCPLETKPLRETRKCIYSIY 176
DB 146 NPCANG-----GQSAFESHYICTCPRNFHGTGRDQVDECAVSPSCRNGCINEV- 198
OY 177 FMKRSVYFICGQO-----CVRTITPRACVLASLAHNAKAPRGVKKALGTASVYMGVN 231
DB 199 -----GSTLCKCPRELYTGRPHCOR-----LYRPLCLSP-----CRSGGTCVQTS-D 237
OY 232 GTGTCCGGLGFNFSTACE-----TCTE-----GRYGIHCD-----QAC----- 263
DB 238 THTTSCSLPFGTGTGOTCHANNVDDCTQHACENGRCIDGINYNCHCDKHWGQYCTEDVDE 297
OY 264 -----SCVHGRCGQPLGSGCDGCVGRGVKCDMEITTDNC-----NGTCH----- 305
DB 298 CELSPNACGNGTCHNTIGGFHCYCVWGTGDCSENI-----DDCSAASCAGATCHDRVAS 355
OY 306 -----TSANCLDP-DGKASCKCAAFRGVYCT----- 334
DB 356 FPECPRHRTGLCHLDACISNPCOKGSCNCDINPVSGKAICTCPRGY--TGSACNODID 413
OY 335 -----AINACTSNGGCKTKA--DCK-----RTTP-----GNR 360
DB 414 ECLSLANPRCEHGRCLNTKSPCKCLQGYEGRGCEMDVNECKSNPCONDATCIDIGGF 473
OY 361 VCVCAGYTGDIIVCLEIN-----PCL-----ENHGG----- 387
DB 474 HCLIMPGY--EGVFC-QINSDDCASQPCLNKCKIDKINSFHCPRKGFSSLCQVDVDEC 530
OY 388 -----CDRNACCTGTGRQAACNCLPKYTGDQKCSL-INVCILNNGGCSFPAPCNYTEQD 442
DB 531 ASPTPCNKAACSTD-GPRKKYTCCECTPGFSGIH--CELDINECAS--SP--CHYGVCR 579
OY 443 ORI-----CTKPRYTGDIYGRGSIYELPKPNSTSOYFPOLOHAYBELAGPRPTVFA 498
DB 580 DGVAFLCQCDRPRGTGR--LCETINI-NECLSQPCRNNGTQODRKNALICICPKG----- 630
OY 499 PLSSSFNHEPRIKDWDQGLMSOVLRYHVVCQQLLDNLKVTTSATTLQGEVVSISVQ 558
DB 631 --TTGVNCEINID-------CKRKCDDYCKIDK----- 656
OY 559 DTYFINNEAVLISDITSTGVHIVDKLLSPKNLLITPPDALGRVYLQNLTTVAANHGYT 618
DB 657 ----INGYECVCEPGYSGSMCNINIDCALPCHNGGTGIDGV-----NSFTCLCPDGF- 706

OY 619 KESKLIODSGLSVYIDSIHPTVFWPTDKALEALPPEQODFLFNODNKDKLSYL--- 675
DB 707 -----RDATCLSGHNCCSSNP-----CIHSCLDQINSYRCVC 739
OY 676 -----KFHYIRDSKALASDLPRASAKTLOGSL-SYRCGTGSDIGELFINE---OMCR 725
DB 740 EAGMMGNRCDININECLSNPCVNGGTCKDMTSGYLCTCRAGFSGPNQMINECASNPC- 798
OY 726 FIHIGLFDYVAAVIGDCLLMNPTLGRCDPFTTFDIPGEC---GSC-----IFPKC 775
DB 799 -LNGSCIDYVAGKCMOML--PTTGEVCEVNLAPCSBPRCKNGCVCRESDPQSFSCNC 855
OY 776 PLKSPKGVK---KKCIYNLPFRRNVEGCONL-----CTVYIOTPRCHGF----- 820
DB 856 PAGMGQTCEDVINECVARNPC---TNGGYCENLNGGFQC-----RCNNGFTGALCEND 905
OY 821 MPDCOACP---GC-----PPTPCNNRMCMDL---Y 845
DB 906 IDDEPRPSCNSGVCCODRVNGFCVCLAGFRGERCAEDIDECVASAPCRNGNCTCVNSY 965
OY 846 TPMGQCLCHGFNGTACEL-----CWHGRGSPD----- 873
DB 966 T---CSPRAGFSGINCETINTPCTESSCFNGTGVGIGISSFCVCLPFTGTANCQHDVN 1021
OY 874 -COPRSCSEHGOODE-----GTTG-----SGEC 895
DB 1022 ECDSPRPNONGSCODDGVGTCTKCTPHGTGLNCQSLRWKDCSSPCKNKGSCMOQGASPTC 1081
OY 896 LCETGWTAAACDPTTAVFANCT--PACSVIATCTENNVCV-----CNLANEG----- 940
DB 1082 QCASGWTGICVDVPSVSGEVAARQGVSAVLCHRHAGQCVADGNTHLRCQAGYTGYSYC 1141
OY 941 -----DGITGVYVDF-----CKQNNNGCAKVAKCSQKGTQV- 971
DB 1142 EQVDECOBPNPCNGATCT--DYLGYSCECVPRGHNMCSEKEIEMCLIS-QPCQNGGCTID 1198
OY 972 -----SCSKKGVYKGDGYS-C-IEIDPCADGV-----NGG-CHEHATCRMTGPGKH 1014
DB 1199 LVNTYKCSQPRGTPQ--GVHCEIDIDDCSPSYDPLTGERPRCNGRCVDRV-----GCV 1249
OY 1015 KCECKSHYVGDVCEPPEBOLPLDRCLDONGCNPDA--CADLYFQDTTVGVPHLRSPFG 1072
DB 1250 GCVCPRAGFVGE--RCEBP--VNECLSD--PCDPSGSYNCVOL-----INDFCECRRTG 1296
OY 1073 ----QYKLTDPKAKEACAKKATATATYNQLSYAKKAYHLCASAMLSGRAYTPTYASQ 1128
DB 1297 YTGKRCETVEFGCKDPTCKNGGTCAVASNTKHGYICK--CQPYSGSS-----CEYDSQ 1348
OY 1129 KCGANVYVIGIDYGSRANKSEMMDVFCYRMKDYNCTCKAGYVGDFSC-----SGNLLQVLM 1184
DB 1349 SCGS-----LRCRN-GATCVSGHLSPPCLCAPRFSGHCOCOTRM 1385
OY 1185 SFPSLTN-----FLTEVLAFSKSSARQ 1207
DB 1386 DSPCLVNPVCYNGTQOPISDAFERYCSCPANFNGLCHILIDYFSSGQGR 1435
RESULT 11
NTC3_MOUSE
ID NTC3_MOUSE STANDARD; PRT; 2318 AA.
AC 061982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch 3 protein.
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathii; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X SWISS WEBSTER;

RX MEDLINE=95001556; PubMed=7918097;
RA Landelli M., Dalstrand J., Landahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium."
RL Mech. Dev. 46:123-136(1994).
CC -1- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC
CC -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC
CC -1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL; X74760; CA52776.1; -.
DR HSSP; P00740; 11XA.
DR MGD; MGI:99460; Notch3.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; ASX_hydroxyl.
DR InterPro: IPR000561; EGF-1like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.
DR Pfam; PF000023; ank; 6.
DR Pfam; PF00008; EGF; 34.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_Ca; 19.
DR SMART; SM00001; EGF_Like; 13.
DR SMART; SM00004; NL; 3.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50297; ANK_REP_REGION; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 18.
DR PROSITE; PS00022; EGF_1; 33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_Ca; 17.
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW ANK repeat; Glycoprotein.
FT DOMAIN 1 1643 EXTRACELLULAR.
FT TRANSMEM 1644 1664 POTENTIAL.
FT DOMAIN 1665 2318 CYTOPLASMIC.
FT DOMAIN 2242 2261 PEST.
FT DOMAIN 39 78 EGF-Like 1.
FT DOMAIN 79 119 EGF-Like 2.
FT DOMAIN 120 157 EGF-Like 3.
FT DOMAIN 159 196 EGF-Like 4.
FT DOMAIN 198 235 EGF-Like 5.
FT DOMAIN 237 273 EGF-Like 6.
FT DOMAIN 275 313 EGF-Like 7.
FT DOMAIN 315 351 EGF-Like 8.
FT DOMAIN 352 390 EGF-Like 9.
FT DOMAIN 392 430 EGF-Like 10.
FT DOMAIN 432 468 EGF-Like 11.
FT DOMAIN 470 506 EGF-Like 12.
FT DOMAIN 508 544 EGF-Like 13.
FT DOMAIN 546 581 EGF-Like 14.
FT DOMAIN 583 619 EGF-Like 15.
FT DOMAIN 621 656 EGF-Like 16.
FT DOMAIN 658 694 EGF-Like 17.
FT DOMAIN 696 731 EGF-Like 18.
FT DOMAIN 735 771 EGF-Like 19.

FT DOMAIN 772 809 EGF-Like 20.
FT DOMAIN 811 848 EGF-Like 21.
FT DOMAIN 850 886 EGF-Like 22.
FT DOMAIN 888 923 EGF-Like 23.
FT DOMAIN 925 961 EGF-Like 24.
FT DOMAIN 963 999 EGF-Like 25.
FT DOMAIN 1001 1035 EGF-Like 26.
FT DOMAIN 1037 1083 EGF-Like 27.
FT DOMAIN 1085 1121 EGF-Like 28.
FT DOMAIN 1123 1159 EGF-Like 29.
FT DOMAIN 1161 1204 EGF-Like 30.
FT DOMAIN 1206 1245 EGF-Like 31.
FT DOMAIN 1247 1288 EGF-Like 32.
FT DOMAIN 1290 1326 EGF-Like 33.
FT DOMAIN 1336 1374 EGF-Like 34.
FT REPEAT 1388 1428 LIN/NOTCH 1.
FT REPEAT 1429 1467 LIN/NOTCH 2.
FT REPEAT 1468 1503 LIN/NOTCH 3.
FT REPEAT 1506 1902 ANK 1.
FT REPEAT 1906 1935 ANK 2.
FT REPEAT 1939 1968 ANK 3.
FT REPEAT 1972 2001 ANK 4.
FT DISULFID 43 55 ANK 5.
FT DISULFID 49 66 BY SIMILARITY.
FT DISULFID 68 77 BY SIMILARITY.
FT DISULFID 83 94 BY SIMILARITY.
FT DISULFID 88 107 BY SIMILARITY.
FT DISULFID 109 118 BY SIMILARITY.
FT DISULFID 124 135 BY SIMILARITY.
FT DISULFID 129 145 BY SIMILARITY.
FT DISULFID 147 156 BY SIMILARITY.
FT DISULFID 163 175 BY SIMILARITY.
FT DISULFID 169 184 BY SIMILARITY.
FT DISULFID 186 195 BY SIMILARITY.
FT DISULFID 202 213 BY SIMILARITY.
FT DISULFID 207 223 BY SIMILARITY.
FT DISULFID 225 234 BY SIMILARITY.
FT DISULFID 241 252 BY SIMILARITY.
FT DISULFID 246 261 BY SIMILARITY.
FT DISULFID 263 272 BY SIMILARITY.
FT DISULFID 279 292 BY SIMILARITY.
FT DISULFID 286 301 BY SIMILARITY.
FT DISULFID 303 312 BY SIMILARITY.
FT DISULFID 319 330 BY SIMILARITY.
FT DISULFID 324 339 BY SIMILARITY.
FT DISULFID 341 350 BY SIMILARITY.
FT DISULFID 356 367 BY SIMILARITY.
FT DISULFID 361 378 BY SIMILARITY.
FT DISULFID 380 389 BY SIMILARITY.
FT DISULFID 396 409 BY SIMILARITY.
FT DISULFID 403 418 BY SIMILARITY.
FT DISULFID 420 429 BY SIMILARITY.
FT DISULFID 436 447 BY SIMILARITY.
FT DISULFID 441 456 BY SIMILARITY.
FT DISULFID 458 467 BY SIMILARITY.
FT DISULFID 474 485 BY SIMILARITY.
FT DISULFID 479 494 BY SIMILARITY.
FT DISULFID 496 505 BY SIMILARITY.
FT DISULFID 512 523 BY SIMILARITY.
FT DISULFID 517 532 BY SIMILARITY.
FT DISULFID 534 543 BY SIMILARITY.
FT DISULFID 550 560 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 571 580 BY SIMILARITY.
FT DISULFID 587 598 BY SIMILARITY.
FT DISULFID 592 607 BY SIMILARITY.
FT DISULFID 609 618 BY SIMILARITY.
FT DISULFID 625 635 BY SIMILARITY.
FT DISULFID 630 644 BY SIMILARITY.
FT DISULFID 646 655 BY SIMILARITY.
FT DISULFID 662 673 BY SIMILARITY.
FT DISULFID 667 682 BY SIMILARITY.

```
FT DISULFD 684 693 BY SIMILARITY.
FT DISULFD 700 710 BY SIMILARITY.
FT DISULFD 705 719 BY SIMILARITY.
FT DISULFD 721 730 BY SIMILARITY.
FT DISULFD 739 750 BY SIMILARITY.
FT DISULFD 744 759 BY SIMILARITY.
FT DISULFD 761 770 BY SIMILARITY.
FT DISULFD 776 787 BY SIMILARITY.
FT DISULFD 781 797 BY SIMILARITY.
FT DISULFD 799 808 BY SIMILARITY.
FT DISULFD 815 827 BY SIMILARITY.
FT DISULFD 821 836 BY SIMILARITY.
FT DISULFD 838 847 BY SIMILARITY.
FT DISULFD 854 865 BY SIMILARITY.
FT DISULFD 859 874 BY SIMILARITY.
FT DISULFD 876 885 BY SIMILARITY.
FT DISULFD 892 902 BY SIMILARITY.
FT DISULFD 897 911 BY SIMILARITY.
FT DISULFD 913 922 BY SIMILARITY.
FT DISULFD 929 940 BY SIMILARITY.
FT DISULFD 934 949 BY SIMILARITY.
FT DISULFD 951 960 BY SIMILARITY.
FT DISULFD 967 978 BY SIMILARITY.
FT DISULFD 972 987 BY SIMILARITY.
FT DISULFD 989 998 BY SIMILARITY.
FT DISULFD 1005 1016 BY SIMILARITY.
FT DISULFD 1010 1023 BY SIMILARITY.
FT DISULFD 1025 1034 BY SIMILARITY.
FT DISULFD 1041 1062 BY SIMILARITY.
FT DISULFD 1056 1071 BY SIMILARITY.
FT DISULFD 1073 1082 BY SIMILARITY.
FT DISULFD 1089 1100 BY SIMILARITY.
FT DISULFD 1094 1109 BY SIMILARITY.
FT DISULFD 1111 1120 BY SIMILARITY.
FT DISULFD 1127 1138 BY SIMILARITY.
FT DISULFD 1132 1147 BY SIMILARITY.
FT DISULFD 1149 1158 BY SIMILARITY.
FT DISULFD 1165 1183 BY SIMILARITY.
FT DISULFD 1177 1192 BY SIMILARITY.
FT DISULFD 1194 1203 BY SIMILARITY.
```

Query Match 6.0%; Score 473; DB 1; Length 2318;
Best local similarity 22.0%; Pred. No. 3.3e-25;
Matches 255; Conservative 101; Mismatches 381; Indels 420; Gaps 70;

```
OY 131 EIQKRCNNNDT-----IIVRGEGCK-CSQAPRPLETKRLRTRKCTIYIYENG 179
DB 280 QIQPNACHNGTGFNMLGSHSCVNGWTGESCSONID-----DCATAVCFHG 327
OY 180 KKSVEFGCPQPCVPTIITPRACWLASIAHNKAPRGVKKMCAIGTASV-----WD 228
DB 328 -----ATGHDRVASFYC-----ACPMKGTGLLCHLDADACVSNPCHEDATCDPN 370
OY 229 GVGNGTGCQCGIGFNGTAC---ETCTEGKYGIHCDQACSVH-GRC--SQGPLDGGCD 281
DB 371 PVSGRATICTCPRGFGACDDVDDECSIG-----ANPCEHLGRCAVNTGCSF---LCO 419
OY 282 CDVGRGVKCOMETITDNC-NGCTHTSANCILDPDGKASKCAAGFRNGIYCTA-IMNC 339
DB 420 GGRGTYGPCEIDV--NEFLSGPCRNQATC-LDRIGQFCICIMAGF--NGTYCEVIDDEC 474
OY 340 ETS---NGCSTKADCKRTTPGNRVCCVCKAGTYTGDIIV-LEINPLENHGCDRNAECT 395
DB 475 OSSPCVNGV-----CKDRVNGFS-CTGSPGFSG--MQGLVDDECAST--PCRNAKCV 524
OY 386 QTPMAVANCILPKYTGDKGVCSLNVCLITJNNGGSPFAFCVNTYDQRI---CTCKPD 451
DB 525 DQ-PDGYECRCAEGFEG-----TLCEARNVDDCSPPD-CHHGRCVDDGIASFSCACABG 574
OY 452 YTGDCIVRGSTYGEIPKPNSTSOYFFQLOEHAVALRGPFTYFAPLSSSFNHPRIK 511
DB 575 YT--GIRCESOV-DECRSOP--CRVGGKCLDLVDYKLCRCRPGT-----TGVNCEVNID 623
```

```
OY 512 DMDQO-----GLMSQ-VLRHYVVCQQL---LDNLKVTTSATTLQGEPSISVSQDTEI 563
DB 624 DCASNCTGTGRCRQDGINRQDV-CQPGFTGPLCNWEINCCASSPCGEG----- 670
OY 564 NNEAKVLSDDIISTNGVHVHIDKLSPKNL--LITPKALGRVLOWLTTVAANHCTKS 621
DB 671 -----GSCVDGENG-FHC---LCPPGSLPPLCLP-----ANH----- 698
OY 622 KLIDQSGLLSVITDSIHTVT-----VFPPIDKALEALPPEQDPLFNG 665
DB 699 -----PCAHKPCSHQVCHDAPGCRVCCEPWSGPRCSQSLAP----- 737
OY 666 DNKDKLSYLFKTHVIRDSKALASDLPRASAMKTLQSELVYRGTSQSDIGELFLNEQMR 725
DB 738 -----ACESQPCQAGTQSDSIGRHCACAPFO----- 766
OY 726 FTHRGLLPVAVYGTIDCLLMPNPLTGGRCDDFTTEDIPECCSGCTPTPKCLSKPKGVK 785
DB 767 -----GHQCEVLSRCPITPLCEH-----GHCESD----- 790
OY 786 KKCITYNPLPERRNVEGQNLCTVYIOTPRCHQYFMPDQACRGCGRDPTCCNNRGMCRDLY 845
DB 791 -----PDRLLTV--CS--CPRGWQPRC-----QQDVDEBAGA--SPCGHGTCTNL- 830
OY 846 TPMG-OCILCHTFNGTACELCWHGRGPRGPCOPRSCSEHGQCEGITGSEGLCETGWTAA 904
DB 831 -FQNFRCIGHRGYTGPRCD-----QDIDCDPRPCLHGGSCQDGV-GFSQSCSLDGFAQR 883
OY 905 SC--DTPTAVFAVCPACGSVHATCTE--NNTQVCNLYEGBDITCTVVDPECKONNGGA 959
DB 884 RCARDVDECLSSBCGP-----GCTDHWASFICACPPGY--GGFHC-IDLPCSSSCF 935
OY 960 KVAKCSQKQTVQSCSKKQYKKGDIYSCIEIDPCADQVNGGCHENATCRMTGPRKHNCEK 1019
DB 936 NGTCVDGVSSFSCLCRPRTYTG-HQYEDAPC---FSRPLCHGICNPTRHG-FECTGR 990
OY 1020 SHYVGDGVCEPEQPLPLDRC---LDQNGCQHPDASCAIYFODTTVGVFHLRSPIGQYK 1075
DB 991 EGFQTS--QC---QNVDMCSQAPCQNGRC----- 1016
OY 1076 LTFDKAKEACAEATITATYTNLSAQKAKYHLCSAGMLESGRV---AYPTTYSQKCG 1131
DB 1017 -----VQTGAYCICPPGM--SGRLDIOQLPCTEAAQMG 1049
OY 1132 ANVVGIVDGSRA-NKSEMWDFCYR-----MKDVCN 1162
DB 1050 VRLEQLCQEGKCIDKGRSHYCVCPREGRTGSHCEHVEDPCTAOPRCQHGCTCKQMGYVC 1109
OY 1163 TCKAGYVGDGFGSCGNL 1179
DB 1110 ECPAGYAGD--SCEDNI 1124
```

RESULT 12
NTCL_RAT STANDARD; PRT: 2531 AA.
AC 007008:
DT 01-NOV-1995 (rel. 32, created)
DT 15-JUL-1999 (rel. 38, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Neurogenic locus notch homolog protein 1 precursor.
GN NOTCH1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Schwann cell;
RX MEDLINE=92111383; PubMed=1764995;
RA Weimaster G., Roberts V.J., Lemke G.;
RT "A homolog of Drosophila Notch expressed during mammalian development.";


```
FT DISULFD 629 638 BY SIMILARITY.
FT DISULFD 645 655 BY SIMILARITY.
FT DISULFD 650 664 BY SIMILARITY.
FT DISULFD 666 675 BY SIMILARITY.
FT DISULFD 682 693 BY SIMILARITY.
FT DISULFD 687 702 BY SIMILARITY.
FT DISULFD 704 713 BY SIMILARITY.
FT DISULFD 720 730 BY SIMILARITY.
FT DISULFD 725 739 BY SIMILARITY.
FT DISULFD 741 750 BY SIMILARITY.
FT DISULFD 757 768 BY SIMILARITY.
FT DISULFD 762 777 BY SIMILARITY.
FT DISULFD 779 788 BY SIMILARITY.
FT DISULFD 795 806 BY SIMILARITY.
FT DISULFD 800 815 BY SIMILARITY.
FT DISULFD 817 826 BY SIMILARITY.
FT DISULFD 833 844 BY SIMILARITY.
FT DISULFD 838 855 BY SIMILARITY.
FT DISULFD 857 866 BY SIMILARITY.
FT DISULFD 873 884 BY SIMILARITY.
FT DISULFD 878 893 BY SIMILARITY.
FT DISULFD 895 904 BY SIMILARITY.
FT DISULFD 911 922 BY SIMILARITY.
FT DISULFD 933 942 BY SIMILARITY.
FT DISULFD 987 998 BY SIMILARITY.
FT DISULFD 992 1007 BY SIMILARITY.
FT DISULFD 1009 1018 BY SIMILARITY.
FT DISULFD 1025 1036 BY SIMILARITY.
FT DISULFD 1030 1045 BY SIMILARITY.
FT DISULFD 1047 1056 BY SIMILARITY.
FT DISULFD 1063 1074 BY SIMILARITY.
FT DISULFD 1068 1083 BY SIMILARITY.
FT DISULFD 1085 1094 BY SIMILARITY.
FT DISULFD 1101 1122 BY SIMILARITY.
FT DISULFD 1116 1131 BY SIMILARITY.
```

Query Match 5.7%; Score 446; DB 1; Length 2531;
Best Local Similarity 21.5%; Pred. No. 3, 1e-23;
Matches 275; Conservative 112; Mismatches 474; Indels 420; Gaps 78;

```
OY 135 NRCNDNTIIVRGCGKSOAPCLETRKREKCIYSIFMGRRS--VFIGCPOQ-- 190
DB 180 NECSQMP-----GLCHRGGTCHNETISYR---CACRATHITGHCELPYPCSPSPC 227
OY 191 -----CYRTITTRACWLASLAHNAKRPARGEVKMKALGTSWMDVNGT 233
DB 228 QNGGTCRPTGDTTHCCACLPGRFAGQNC-----EENVDDCPG--NNCKNGGACV-DGVN-T 278
OY 234 GTCOCGLGFNG-----TACE---TCTEGKYGJHC-----DQACS 264
DB 279 YNCRCPREPTGQYCTIEDVEECOLMPNACONAGTCHNSHGVCVNGMTGEDCSNID 338
OY 265 CVHGRCQGP-----LGDSCDCDVGWGVKCDMEITTDNC--NGTCHTSANCLLDP-DGK 317
DB 339 CASACFOGATGCHDRYASVYCEQPHGRTLLCHL---NNAJCSNPENESNCDOTNVNCK 395
OY 318 ASCKCAAGRCNGYCTA--INACETISNGGCKTADCKRTTPGNGRVGVCAAGYTGDIYVC- 375
DB 396 AICTCPRGY--TGPAQSQVVDICALGAPCEHAGKCLNTL-GSFECOCLOGYGR--CE 450
OY 376 LEINPLENHGCGDRNAEC--TGTGPNQAVCNCLPKYTGSGKVCSTL-----INVCCLNNG 428
DB 451 IDVNECISN--FCQNDATLIDGIGFO--CLICMPG--BGVYCEINTDECASPCL-HNG 503
OY 429 GCSPAFCVYTEDODRICCTCKPDYTGDIYVCRGSIVGELPKPNSTSOYEFQLOEHAVREL 488
DB 504 RC-----VDKINEFLCCCPKFGSGH--LCQYDV-DECASTRPC-----KNAKACL 544
OY 489 AGRGPTTVAFLSSSFNN-EPRKMD-----QOGLM----- 519
DB 545 DGRNNTYTCVCTEGYTGTHCEVDIDECDDPDCHIGLCKDGVAFPTCLCPGVTGHHNCEINI 604
```

```
OY 520 ----SQVLRYHYVCCOQLLDN-----LKVTT-----SATTLQGEVPSISVSDTVEI 563
DB 605 NECHSQPCR-HGGTQCD--RDNYLTLCLCKGTTPGRCNCELINDCCASNPCDSSGCLDKI- 659
OY 564 NNEAKVLSSDIISTNGVIVH-IDKLL-SPKNLLITPKDALGRVLOMLTYVAANHGYTKRS 621
DB 660 --DGYECACPEYTGSMCVNVIDECAGSPCHNGTCEDEGIAGR-----TCRCEGY----- 708
OY 622 KLIQDSGLISVITDSIHTVYTVWPTDKALEALPRQDQFLFNQDKDKLSYLKHHVIR 681
DB 709 ---HDPCTCSEVNECSNP-----CLHGACRGDINGY----- 737
OY 682 DSKALASDI.PRSASWKTLOGSELSVRCGTGSDIGELFLNEQWCRFTIHRGLLFDVGAAYGI 741
DB 738 -----KDCDCAIGWS-----GTNCDLINNECESNIPC--VNGGICKD--MTSGV 775
OY 742 DCLLMNPITLQRCDFPTTFDIPGEC--GSCI-----FPPKCPIL----- 777
DB 776 VCTCRGEGFSGPNCQTINI.ECASNPCINOGTCIDVAGYKNCNPLPYTGATCEVVLAPCAT 835
OY 778 -KSKPRGVKK-----KCIYNPLPRRNVBGCN----- 804
DB 836 SPCKNSGVCKESEDYESFSCVCPFGWOGOTCEIDINECVKSPC--RHGASCONTNGSYR 892
OY 805 -LCTTVIQTTPRC-----CH-----GYFMPDCOACPG-----GPD 833
DB 893 CLCOAGYTGGRNCSDDIDDCRPNCHNGSGCTDGVNNAFCDCILPFGPGATCEEDIINECATN 952
OY 834 PCNNRGMCRDL---YTPMGOCILHTGFNGTACELCMHGRFGPDCQPRSGSEHGOCDEGIT 890
DB 953 PCONGANCTDCVDSYF---CTCPTGFNGIHC-----NNTPDCTESSCFNGGTGVDGI- 1002
OY 891 GSGECICETGWTASCDTTPAVFAVCTPACSVIATCTEN---NTVCNUNYBEDGITT- 946
DB 1003 NSFTCLCPREFGTGYCYQYD--VNECDSPCLHGCTQDSTGYTKCPCPGY--TGLNCON 1058
OY 947 VVDFCKONNGGCAKAVKCSQKGTQVSCSKKGYGKGDYSCIEIDPCADGVNGSCH----- 1001
DB 1059 LVNWC--DSAPCKNGKCMQNTNQYHCECRSG--TGPNK-----DVLVSCEVAAGK 1107
OY 1002 -----EHATCRMTGPKKIKCECKSHYVGWVDCERBQLPDLKCLQDNQCHPDASC 1052
DB 1108 RGIQVTLLOHGGLCYDEEDKHCHCOAGYTG--YCEDE---VDEB--SPNQCQMGATC 1160
OY 1053 ADLYFODTIVGVPHLSPGLQYKILTFDKAKKACAKAATITATYNOLSYQAKAYHILCSAG 1112
DB 1161 TDY-----LGRFSCKCVAGYHGSNCSEETNECLSDPC---ONGGTCLDLTTPYKCSG- 1209
OY 1113 WLESGRVAYPTTYASQKCGANVGI---VDYGRANKSEKMDVFCY-----RMKDVNC 1162
DB 1210 -----PRGIQVYHCELINVDCHPRLDPASRSK-----CFNNGTCVVDVGGYTC 1253
OY 1163 TCKAGYVGDGFGSCSGNLOVL 1183
DB 1254 TCRPGFVGE--RCGDVNECL 1272
```

RESULT 13
NTCL_MOUSE
AC 001705;
ID NTCL_MOUSE STANDARD; PRT: 2531 AA.
DT 01-NOV-1995 (rel. 32, created)
DT 01-FEB-1996 (rel. 33, last sequence update)
DT 16-OCT-2001 (rel. 40, last annotation update)
DE Neutrogenic locus notch homolog protein 1 precursor (NOTCH protein).
GN NOTCH1 OR MOTCH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN 111
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;

ET DISULFID 532 543 BY SIMILARITY.
ET DISULFID 537 552 BY SIMILARITY.
ET DISULFID 554 563 BY SIMILARITY.
ET DISULFID 570 580 BY SIMILARITY.
ET DISULFID 575 589 BY SIMILARITY.
ET DISULFID 591 600 BY SIMILARITY.
ET DISULFID 607 618 BY SIMILARITY.
ET DISULFID 612 627 BY SIMILARITY.
ET DISULFID 629 638 BY SIMILARITY.
ET DISULFID 645 655 BY SIMILARITY.
ET DISULFID 650 664 BY SIMILARITY.
ET DISULFID 666 675 BY SIMILARITY.
ET DISULFID 682 693 BY SIMILARITY.
ET DISULFID 687 702 BY SIMILARITY.
ET DISULFID 704 713 BY SIMILARITY.
ET DISULFID 720 730 BY SIMILARITY.
ET DISULFID 725 739 BY SIMILARITY.
ET DISULFID 741 750 BY SIMILARITY.
ET DISULFID 757 768 BY SIMILARITY.
ET DISULFID 762 777 BY SIMILARITY.
ET DISULFID 779 788 BY SIMILARITY.
ET DISULFID 795 806 BY SIMILARITY.
ET DISULFID 800 815 BY SIMILARITY.
ET DISULFID 817 826 BY SIMILARITY.
ET DISULFID 833 844 BY SIMILARITY.
ET DISULFID 838 855 BY SIMILARITY.
ET DISULFID 857 866 BY SIMILARITY.
ET DISULFID 873 884 BY SIMILARITY.
ET DISULFID 878 893 BY SIMILARITY.
ET DISULFID 895 904 BY SIMILARITY.
ET DISULFID 911 922 BY SIMILARITY.
ET DISULFID 916 931 BY SIMILARITY.
ET DISULFID 933 942 BY SIMILARITY.
ET DISULFID 987 998 BY SIMILARITY.
ET DISULFID 992 1007 BY SIMILARITY.
ET DISULFID 1009 1018 BY SIMILARITY.
ET DISULFID 1025 1036 BY SIMILARITY.
ET DISULFID 1030 1045 BY SIMILARITY.
ET DISULFID 1047 1056 BY SIMILARITY.
ET DISULFID 1063 1074 BY SIMILARITY.

Query Match 5.7%; Score 445; DB 1; Length 2531;
Best Local Similarity 22.6%; Pred. No. 3.7e-23;
Matches 288; Conservative 104; Mismatches 477; Indels 404; Gaps 84;

OY 135 NCNDNDITIVGECGKCSQO-----APCPLETKPLRTKRCYISYFMGKRSVFTGCO 188
DB 180 NCCSONPFGICRHG--GHCHNEIGSYRCACCATHTGPHCEL-----PYVPCS 223
OY 189 PC-CVRTIITR-----ACMLASLA-----HNKKPAPGEVKMCALGTASVMDGVNGT 233
DB 224 PEPCCMGACRTGTGTTTHCAC-LGCFAGCNCEENVDCPG--NNCKNGACY-DGVN-T 278
OY 234 GTCCGGLGNG-----TACE--TC--TEGKYIHC-----DQACS----- 264
DB 279 YNCRBPPEVTGYCTEDVDECOLMPACONAGCHNTHGYNCOVYNGWTGECSENID 338
OY 265 CYHGKCSQGP-----LGDSGCCDVGMRGVKCDMETTTNCGTCHTSANCLIDP-DGKA 318
DB 339 CASACFCQATGCHDRVAFSECPHGRTGLCHLKHAC--ISNPCEGSCNCDTNPVNGKR 396
OY 319 SCKAAGFNGGTVCYA-INACETSNNGGCTKADCKRTTGNRNVCVCKAGYTGDIIVC-L 376
DB 397 ICTCSBG--TGPACSDYVDECDLGNRCENHAGKCLNTL-GSEFCCLOGYTPG--CEI 451
OY 377 EINPLENHGGCDRNAEC-TGTGPNDAVNCCLPKYTGDKVSL-----INVCLTNNGG 429
DB 452 DVNECISN--PCNDATCLDQIGERQ--CICMPGY--EGVYCEINTDECASSPCL-HNH 504
OY 430 C-----SPPAF-----CNYTEQD-----ORICTCKPDYTGDIIV 458
DB 505 CMDKIHFOCQCPKGFNGHLCQYDVDECASTPCKNGAKCLDGPNTYTCVCTEGYTGTH-- 562

OY 459 CRGSIVGELPKNESTSYQFFLOEHAVERLAGPPTVEAPLSSPNH-EPRIKDWDOG 517
DB 563 CEVDI-DECDPDPG-----HGSCKDGAFTTCLQOQRYTGHHEITNNE----- 606
OY 518 LMSQVLRHYVVGCOQLLDN-----LKVTT-----SATTLQGPVSVISVODTYINN 565
DB 607 CHSQPCR-HGGTCQD--RDNSTYLCCLKGTGTGENCEININDCASNPDCSDGTCLDKI---- 659
OY 566 EAKVLSDDIISNGVIAH-IDKLL-SPKNLLITPKALGRVLONLITVAANHGYTFSKL 623
DB 660 DGEYCEACEPGYGVGMCNVNIDECAGSPCHNGGTCEGDIAGF-----TCRCPEGY----- 708
OY 624 IDDSGLSVITSTSIHPPVIVFEMPTDKALPDPQODPLFNQDNKDKLSYLFKHFVTRDS 683
DB 709 -HDPTCLSEVNECSNP-----CIRHACRGLNGY----- 737
OY 684 KALASDLPRASAKTKLQGSSELSVRCGTSGDISELFLNEQMCREFIRHGLLFYDVAVYIDC 743
DB 738 -----KCDCAPGWS-----GTNCDINNNECESNDC--VNGGTCKD--MTSGYVC 777
OY 744 LLMNPTLGRCDTFTTFDILPGEC--GSCI-----FTPKCPL-----K 778
DB 778 TCREGFSFGNCCNTINNECASNPCLNCGTCTIDVAGYKCNCPLYTGATCEVYLAICATSP 837
OY 779 SKPKGVK-----KCIYNPLFPRRVNCECON-----L 805
DB 838 CKNSGVCKSEDEYESTSCVCPICMOGQOCEVDINECVKSPC--RHGACQOINTNGSYRL 894
OY 806 CTIVVIQTPRC-----CH-----GYFMPDQACPG-----GPDPC 835
DB 895 COAGYGRNCESDIDDCRPNCHNGGCTDINTAFCDCLPGFGAFCEEDINECASNP 954
OY 836 NNRGMKRD---YTPMGQCLHTGFNGTACELCMHGRFPRDQPRGCSHGGCDEBITGS 892
DB 955 QNGANCTDQVDSY---CTCPVGFNGHCE-----NNTPRCTESSCFNGGTCVDDI-NS 1004
OY 893 GECICETGTAASCDPTPAVFAVCTPACSVHATCTEN--NTCVCNLINEGDOITC--TVV 948
DB 1005 FTICLCPGTGTSQCYD--VNECDSRPLHGTGQDSYGTCTCPQGI--TGLMCONLY 1060
OY 949 DFCCKNNGGCAVAKCSQKGTQVSCSKKRYKGD-----GYSCIEIDPCADGVNGG--CH 1001
DB 1061 RMC--DSAPCKNNGRCWMQNTQYHCECRBSGWTGVNCDVLSVSC-EVAAGKRGIDVLLQ 1117
OY 1002 EHATCMTGPKHKKCEKSKHYGVGDYDCEPEBOLPLDRCLQDNGQCHPRDASCADLYFDPTT 1061
DB 1118 HGLICVDEG-DKHVHCQAGYGTG--YCEDE--VDEC--SPNPCCNGATCTDY----- 1163
OY 1062 VGVFPHLRSPILGOY-KLTFPKAKECAKKAATATATVQLSYAOKAKYHLCAGMLESGRYA 1120
DB 1164 LGFSCSKVAGTHGNSCSEINECISQPCONGGTCTIDLINSYK-----CSC----- 1209
OY 1121 YPTTVAOSKCGANVGI---VDYGRANKSEMDVFCY-----RMKDVCNCTCKAGYV 1170
DB 1210 -PRGTGVHCEIINVDCHRPDPASRSPK-----CFNNGTCVDPDVGVTCTCPPGFVG 1261
OY 1171 DGFSCGNLLQYL 1183
DB 1262 E--RCEGDVNECL 1272

RESULT 14
NTC4_MOUSE
AC P31695; 062389;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurogenic locus notch homolog protein 4 precursor (Transforming protein INT-3).
GN NOTCH4 OR INT3 OR INT-3.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

FT	DISULFID	514	525	BY SIMILARITY.
FT	DISULFID	519	534	BY SIMILARITY.
FT	DISULFID	536	545	BY SIMILARITY.
FT	DISULFID	552	563	BY SIMILARITY.
FT	DISULFID	557	572	BY SIMILARITY.
FT	DISULFID	574	583	BY SIMILARITY.
FT	DISULFID	590	601	BY SIMILARITY.
FT	DISULFID	595	610	BY SIMILARITY.
FT	DISULFID	612	621	BY SIMILARITY.
FT	DISULFID	626	637	BY SIMILARITY.
FT	DISULFID	631	646	BY SIMILARITY.
FT	DISULFID	648	655	BY SIMILARITY.
FT	DISULFID	662	669	BY SIMILARITY.
FT	DISULFID	664	674	BY SIMILARITY.
FT	DISULFID	676	685	BY SIMILARITY.
FT	DISULFID	692	703	BY SIMILARITY.
FT	DISULFID	697	712	BY SIMILARITY.
FT	DISULFID	714	723	BY SIMILARITY.
FT	DISULFID	730	741	BY SIMILARITY.
FT	DISULFID	735	750	BY SIMILARITY.
FT	DISULFID	752	761	BY SIMILARITY.
FT	DISULFID	768	779	BY SIMILARITY.
FT	DISULFID	773	788	BY SIMILARITY.
FT	DISULFID	790	799	BY SIMILARITY.
FT	DISULFID	807	818	BY SIMILARITY.
FT	DISULFID	812	827	BY SIMILARITY.
FT	DISULFID	829	838	BY SIMILARITY.
FT	DISULFID	845	856	BY SIMILARITY.
FT	DISULFID	850	865	BY SIMILARITY.
FT	DISULFID	867	876	BY SIMILARITY.
FT	DISULFID	882	903	BY SIMILARITY.
FT	DISULFID	897	912	BY SIMILARITY.
FT	DISULFID	914	923	BY SIMILARITY.
FT	DISULFID	930	941	BY SIMILARITY.
FT	DISULFID	935	950	BY SIMILARITY.
FT	DISULFID	952	961	BY SIMILARITY.
FT	DISULFID	968	979	BY SIMILARITY.
FT	DISULFID	973	988	BY SIMILARITY.
FT	DISULFID	990	999	BY SIMILARITY.
FT	DISULFID	1006	1019	BY SIMILARITY.
FT	DISULFID	1011	1028	BY SIMILARITY.
FT	DISULFID	1030	1039	BY SIMILARITY.
FT	DISULFID	1046	1057	BY SIMILARITY.
FT	DISULFID	1051	1069	BY SIMILARITY.

```

Query Match 5.6%; Score 436.5; DB 1; length 1964;
Best Local Similarity 21.8%; Pred. 0.11e-22;
Matches 226; Conservative 95; Mismatches 347; Indels 371; Gaps 64.

OY 187 GCP-----QCVVITITRACWL-----ASLHNKK-----PAPGEYKCAL--GTAS 225
Db 179 CRRPGEGHNTCEBDI--NRCFILEPGCRPGTSCNHTNLGSIQCLCPVGOBEPCKLRKACPR 236
OY 226 VWDGVNGTGTCTC-----CGLGFNGTACE-----TCTEG--KY 255
Db 237 PGCSLNG--GTCOLVPEGHSTFHLCLCPGFTGLDEEMNPDCVRIHQCGANGATCLDGLDTRY 295
OY 256 GIHQDQA-----CS-----CYHGKCSQGPLGDSGDCDDCVHGRKYKCMETIT 297
Db 296 TCLCRKTYMKGWDSCEDIDCECARGPERRCKNGTGTCNTAGSFHCVCYSGGAGCEBNTL-- 353
OY 298 DNC--NGTCHTSANCLLDPDGKASCKCAAGFRNGTVCTA INNCETNSNGGCSKADCKRTT 356
Db 354 DDCAATATAPGSTC--IDRVGSFSCLCPRP--RTGLLCHIEDMCLSO--PCHYNAOCS--TN 407
OY 357 P--GNRVCVCVCAKYTGSDIVC--LETINPCLFNHG--CDRNAECTQGTGNAQVACNCLPKY 410
Db 408 PLTGSLTLCICPGYSGS--TCHQDIDDECOMAQGQSPCEHGSGCINT--PGSFNCLCLPKY 464
OY 411 TG-----DGKVCSL--INVCLTINNGGCSFP 433
Db 465 TGSRCADHNECLSPCHPGSTCLDLATFNHCLCPGLGELGICEVEVMDGTSN--PCLNO 522

```

[illegible]

FT	CARBOHYD	3665	3665	N-LINKED (GICNAC. .) (POTENTIAL).
FT	CARBOHYD	4140	4140	N-LINKED (GICNAC. .) (POTENTIAL).
FT	VARSPLIC	1	3616	MISSING (1IN ISOFORM XB-SHORT).
FT	CONFLICT	135	135	G -> G (1IN REF. 2).
FT	CONFLICT	4038	4038	P -> G (1IN REF. 3 AND 5).
FT	CONFLICT	4163	4163	M -> I (1IN REF. 3 AND 5).
FO	SEQUENCE	4289	AA: 464454	MM: 846576126556470
				CM64:

Query Match	5.58;	Score 435;	DB 1;	Length 4289;
Best Local Similarity	23.18;	Pred. No. 3.8e-22;		
Matches 223;	Conservative 71;	Mismatches 318;	Indels 354;	Gaps 58;

Search completed: June 17, 2002, 12:35:32
Job time: 308 sec

```

OY 1047 HPDMSCA--DLIEPDFTVVGFIHLRSPGLGYKLLFPDKAKENC--AKENAAITATYNOLSYAQ 1102
Db 730 H-DQSCVCKDGY-----AGEDCGEARVPSASATDQGLAP 764
OY 1103 KAKYHL 1108
Db 765 GOEAYOV 770

```

QY	186	GCQCGVATITTRACVMASTAHNAKRAPGEV---KKCALGTASVMBGVNGTGTCCGCGLGE	242
Db	116	GLKRGCC-----TGCCCAAS-----AQAAGTGDTVTRTLCSL---HAYED---LSRCTGCSCEPQM	162
QY	243	NGTACETCTE-----GKYGICHDQACSVYHGRSGOGLDGSDDCDVGMGRGVK	290
Db	163	GGPFCSPDPTAEIIRPSSPSPASGSCPPDDCNDGRCVYGRGV-----CFPGYGPS	212
QY	291	CDMETTDNCGNGTCHTSANCLLDPDGKASCKCAAGFRGNGVCTALINACETISNGGCSSTKA	350
Db	213	CGW-----PSCGCGDQGRGRGV-----QGVYVCRAEP---SGPPDSO---RSCPRA---GCSQSG	257
QY	351	DCRRTTPGNRNCVCAKAGYTDGIVCILEINPCLNHHGSCBMAEACTGTGNOAVCMCLPRY	410
Db	258	RCE-----GGR---CYVDDPRTYTD---CGMRSCPR---GCSQGRGC-----ENGRVCYNBYG	301
QY	411	TGDGKVCSLINVCLTNNGGCSPPAFNCYTEQDRICTCKPDTYTDGIVGRGSIYELPKN	470
Db	302	TGED---CG-VRSQ---PRGCSORGR---KDRG---CVDDPBYTGEDCGTSGC-----	342
QY	471	PSTQYFFQJQEHAVRBLACRPFYFAVPLSSFFNIHPRIKIMDDOGGLMSOYLRYHNVYC	530
Db	343	-----PMD-----CQE	348
QY	531	QQLLDNLKYTTSATTLQGEPSVISVSQDVFINNEAKYLLSDIISTNGVIHVIDKLSP	590
Db	349	GGRCVVDGRCVCMPIYT---GEDCSHRT-----	372
QY	591	KNLLITRKDALGRVLOWLTVVAANHGYTKRSKLQDSGLSYTTD---SIHTFYVYW	645
Db	373	-----CPRDCKGRRCEDGECICTDTGYSG-----DDCGVRSCTGDCDNOGRGEDGRGVCM	422
QY	646	P-----TKALALELPFGQDFLFNODNNDKLSYLFKPHVIRDSKALASDLPRASMKTLQGG	701
Db	423	PGYGTGDCGSAC-----RDCRGRG-----	443
QY	702	SELSVRCGTSGSIDIGELFLNBOMCRFIHRLGLFDVGV---AVGIDCLLMPNPLTLAGRCDTFTTF	760
Db	444	-----RCENGVCV-----CNAGYSG---EDCGVRSCTGDC-----BGRGRCESGRCM	482
QY	761	DIREBCGSCIFTTRCPKLAKSRPKG---YKKCIATNP-----LPRRANVEGQ---NLCTVVI	810
Db	483	CWPGYTGDCGTRACPDCKGRGCVYDGRVCVCMPIGTGTEGDCSGRRCPGDRKHGICDEGV	542
QY	811	QTPRCCHGYFMPDC---OACRPGRPDPCNNRGMCRDLYTPMGQCLCTHTGNGTACELCWHG	868
Db	543	CV---CDAGYSEDECSSTRSCGG---CRGRGQGLD-----GRVVCIDYGSGEDC-----	585
QY	869	RFGPDCCPRGSEHNGQCDDEGTTGSGBELCTGTCWTAASCDTPPAVFAVCTPRACSVNATCTE	928
Db	586	--GVRCQPNDCSQHGVCODEV---CICMEGVASEDCSIRT-----CPSNCHGRGR---E	632
QY	929	NNNTVCNLTNEGDDITCTVVDFCFCONNGGCAKAVAKSOKG---TOVSCSSCKKYGKGGYSC	986
Db	633	EGRCRLCPBGTGP---TC-ATRMCP-----ADCRGRRCVQSVCLCHVYGGE---DC	678
QY	987	TEIDPCADGVNCGCHENATCRMYTEPKCHNKCECKSHYVGDVDCEPQRLPLBRLCLODNGOC	1046
Db	679	GOEPPRPSACVGGCGGPRELIR---AGQCVCEGG---KRPDALIOTGCD---CRGRGEC	729

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:25:54 : Search time 19.98 Seconds
(without alignments)
1749.401 Million cell updates/sec

Title: US-09-842-930A-2

Perfect score: 7861

Sequence: 1 SLPSLTLEQMPDYSIFRG.....WGHGPDMSQATTVTPR 1431

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued_Patents_AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backliles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	486	6.2	2523	1 US-08-185-432-18	Sequence 18, Appl
2	480.5	6.1	2556	1 US-08-083-590A-20	Sequence 20, Appl
3	480.5	6.1	2556	3 US-08-532-384-20	Sequence 20, Appl
4	475	6.0	2703	1 US-08-185-432-19	Sequence 19, Appl
5	471.5	6.0	2556	1 US-08-185-432-17	Sequence 17, Appl
6	450.5	5.7	2471	1 US-08-185-432-16	Sequence 16, Appl
7	450.5	5.7	2471	1 US-08-083-590A-19	Sequence 19, Appl
8	450.5	5.7	2471	3 US-08-532-384-19	Sequence 19, Appl
9	413.5	5.3	1404	2 US-08-400-159-2	Sequence 2, Appl
10	413.5	5.3	1404	2 US-08-611-729A-2	Sequence 2, Appl
11	410	5.2	1193	2 US-08-400-159-10	Sequence 10, Appl
12	410	5.2	1193	2 US-08-611-729A-10	Sequence 10, Appl
13	399.5	5.1	1010	4 US-08-882-046-7	Sequence 7, Appl
14	399.5	5.1	1036	4 US-09-068-740A-6	Sequence 6, Appl
15	399.5	5.1	1187	4 US-09-068-740A-7	Sequence 7, Appl
16	399.5	5.1	1218	4 US-08-400-159-6	Sequence 6, Appl
17	399.5	5.1	1218	3 US-08-611-729A-6	Sequence 6, Appl
18	399.5	5.1	1218	4 US-08-882-046-2	Sequence 2, Appl
19	399.5	5.1	1218	4 US-09-214-278-7	Sequence 7, Appl
20	399.5	5.1	1218	4 US-09-068-740A-11	Sequence 11, Appl
21	392	5.0	1219	4 US-08-882-046-5	Sequence 5, Appl
22	368	4.7	3111	2 US-08-125-077-4	Sequence 4, Appl
23	368	4.7	3111	2 US-08-125-077-4	Sequence 4, Appl
24	348	4.4	4544	1 US-08-469-486-52	Sequence 52, Appl
25	348	4.4	4544	2 US-08-469-486-52	Sequence 52, Appl
26	343	4.4	1148	4 US-08-882-046-4	Sequence 4, Appl
27	342.5	4.4	1065	2 US-08-400-159-8	Sequence 8, Appl

28	341.5	4.3	1055	4 US-09-214-278-2	Sequence 2, Appl
29	341.5	4.3	1212	4 US-09-214-278-3	Sequence 3, Appl
30	341.5	4.3	1257	3 US-08-611-729A-8	Sequence 8, Appl
31	339.5	4.3	1248	4 US-08-882-046-6	Sequence 6, Appl
32	337.5	4.3	3075	2 US-08-460-309-5	Sequence 5, Appl
33	337.5	4.3	3075	2 US-08-125-077-5	Sequence 5, Appl
34	331	4.2	1238	4 US-09-214-278-5	Sequence 5, Appl
35	325	4.1	2199	5 PCT-US95-11684-2	Sequence 2, Appl
36	324.5	4.1	1810	5 PCT-US95-11684-4	Sequence 4, Appl
37	322	4.1	1525	3 US-09-191-647-2	Sequence 2, Appl
38	322	4.1	1525	3 US-09-540-245A-2	Sequence 2, Appl
39	322	4.1	1525	4 US-09-540-153-2	Sequence 2, Appl
40	279	3.5	713	3 US-08-872-855-5	Sequence 5, Appl
41	278.5	3.5	830	3 US-08-872-855-11	Sequence 11, Appl
42	276.5	3.5	833	1 US-08-264-534-6	Sequence 6, Appl
43	276.5	3.5	833	1 US-08-083-590A-2	Sequence 2, Appl
44	276.5	3.5	833	1 US-08-465-500-6	Sequence 6, Appl
45	276.5	3.5	833	2 US-08-346-126-6	Sequence 6, Appl

ALIGNMENTS

```
RESULT 1
US-08-185-432-18
; Sequence 18, Application US/08185432
; Patent No. 5750652
;
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Busseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DELTEx PROTEINS, NUCLEIC ACIDS, AND
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/185,432
; FILING DATE: 21-JAN-1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Mistock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
;
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2523 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
;
; US-08-185-432-18
```

Query Match 6.2% Score 486; DB 1: Length 2523;
Best Local Similarity 21.9%; Pred. No. 86-31;
Matches 302; Conservative 131; Mismatches 457; Indels 486; Gaps 93;

OY	105	YVNEAPINIV-----	ANATKGVIHGLEKLEIOLKNRDNNDIIYREGG	150	
Dd	85	FICHCPGFDFDKVCLLPVDNACVANNPCRRNGTICELLNSTE-YKCRCPGMT----	GD--	137	
OY	151	KCSQOAFPLEPFRPLRETRKCI-YSIYFMGRKSVPITGCP-----		189	
Dd	138	SCQDADPC--ASNPNCANGGCCLPFELIQIYCK-----	CPRGHGATCKODINECSQNPCK	189	
OY	190	---OCVFTTIRACWLAS--LAHN-----	AKPADEVKKCALGTASWDGVNTGTICO	237	
Dd	190	NGGCCJINEFSGSYRTCONFRTPGRNCDEPYVPCNSP-----	C-----LNG-GTCR	233	
OY	238	-----CGIGFNPTAGE-----	TCTEG-KYGIHC-----	263	
Dd	234	QTDDTSYDTCTLPEFSQCNCEENIDCPSNNCRNGTGYDVGWTTVMYCOPRPMTOGYLTE	293		
OY	264	-----SCVHRCSOGPLDGSDCDGVNRGVKCMETITTDNC-NGTCHTSANCLL	312		
Dd	294	DVECOQLMPNACQMGCHNTGYGVNVCVNMGTGEDCSBNI--DDCANAAHGSAITC--	349		
OY	313	DPDGKAS--CKCAAGFGNGTVCTAINACTSNNGGSTAADCRRTP--GNRYCVCKAGY	368		
Dd	350	-HDRVASFYECBPBG--RTGLILCHLNAC--ISNPNEGSSNCD-TBVGAKALICTCPXY	403		
OY	369	TG-----DLIYC-LEINPCLNEHGGCORNAECTOT--GPNAVONACLPKTYTGOKVCSL-I	420		
Dd	404	TGPACNNDVDECSTGANPC--EHGG-----RCNTITLASFQ--CNCQGYTAGRP--CEIDY	452		
OY	421	NVCLTNNGGSPPFAFCVYTEDODRICTKPDYTGDIYVGRSITYGELPKNP-----	471		
Dd	453	NECLSN--PCQNDBTC-LDJIGEHOCTCMHGY--EGLYCTNI-DECASNPCLHNGKCID	506		
OY	472	-----STSQYFFOLOEHAIVR-----	LARGPPTYAPLSSSFNH-----	506	
Dd	507	KINEFRDCDPTGEFSGNLCOHDPECTSTPCKNGAKLJDGPSNSTCCOTBEFTGRHEODI	566		
OY	507	EPRIKMDQOGLMSOVLRHYHVGQ-----OLLLDNLKVTTSTLTLOGEVVS-----	I	554	
Dd	567	NECIPDPCHGTCKDGIATFTCTLCRPYTRRLCDNDINECLSKPCLANGGOCTDRENGYIC	626		
OY	555	SVSODTFVINNEAKV--LSSDIISTNGVHVID-----	KLSPKRULLI-----	595	
Dd	627	TCPRKTGVNCEMTIIDOCASMLONGKCIDKIDSBECTCEPGVTGKLC--NINIIEGDS	683		
OY	596	-----TPKDALGRVLONTTVAANHGYRKFSKLIDSGLSLYITD-----	SIHT	639	
Dd	684	NPCRNGSGCKDOI-----NGFTVCPRDGY-----	HDMHCLSEVNECNSNPCIHGACHD	731	
OY	640	PVTYV-----WPTRKALEALPREQODPLFENDONKRLKSLYKPHYIRBSPLASDLBR	692		
Dd	732	GYNKGKCDCEAGW-----	SGSNC-----INNNEESNPCAN	763	
OY	693	SASMKTLOGSELASYRCGTG-----SDIGELFLNEOMCFRIHRGLFDVGAVYGIDCLL	745		
Dd	764	GGTCKDMIGAICYI-CRAGFSGRCNQININECSSNP--C-LMHGTCIDOVAKYKCNCML	818		
OY	746	-----MNPLL-----	GGRC-----DFTTFPI--PG-----	ECGSCIFTP--	773
Dd	819	PYTGAIICEAVLIAPCAGSPCKNGSRCKESEDEFEFSECPCPPMOWGONCEIDMNCVNRPQR	878		
OY	774	-----KPRLSKPRGVKKKCIYNLPRRRAVBECQNCITVYIOTPRGCH----		817	
Dd	879	NGATQNTNGSYKC--NCKPgyTRMC-----	EMIDIOQ-----	PWCCHGGS	920
OY	818	----GYFMPDQACAPRG-----	PDTPCNNNGMKRDLL--YTPMGOLCLHTGFN	858	
Dd	921	CSDGJINMFECN-CRPGFEPGPCCEEDINECASNPCKNKAANTCDVNSTT-----	CTCQGRFS	975	
OY	859	GTACHELWHGRRFPDPOCRSRCSEHQOCDDEGITSGECLJCETGWTAAASCDTPYAVAYCTP	918		
Dd	976	GIHCE-----SNTPRCJTSSCFNGSTCIDGI-NFTQCOPPGFTGSYCGQHD--INECDK	1022		
OY	919	ACSVAHTTEN-----NTVCYNINTEBDGITC-TVVDFCKONNGGCAKAKAKSOkgTOVSCS	974		

Db	1028	PCJNGGTCDDSYGTYKCTCPGY--TGJNCOMJLVHMC--DSSPCKNGKCKMOTNNFYCE	1083
Qy	975	CKRGYNG-----DGYSCIFEDPCADGVN--GGCHEATCRMYGPKHKCEKSHVGVGV	1027
Db	1084	CKSGWTGTYCVDPSVSC-EVAAKQCGVDLVHLCRSMGCVDTGG-NTHRCRQAGTGS-	1139
Qy	1028	DCEPOLLPLDRLQDNGQCHPPDASCADLYFDPTTVGVFLHSLPLQGYKLTEDKAKKACAC	1087
Db	1140	YCEEQ-----VDEC--SPNCPNGATCTDY-----LGGY-----	1166
Qy	1088	EAATITATYNOLSYAOKA---YHLCSAG-----WESGRVAYPTTYVASOKGANYGVIVD	1139
Db	1167	SCBEVAGHGVCNSEEINECLSHPCQNGSTCIDLINTYKSCSPRGTOGVHEINVDCTP	1226
Qy	1140	YGSRAKNSKMWDF-----CY-----RMKVQNCCKRGYGVDFSGNLQVLT	1183
Db	1227	F-----YDSFTLEPKCFNNKCGKIDWGVGYNCLCPGPFVGE--KEGVDNCEL	1271

```

RESULT      2
US-08-083-590A-20
: Sequence 20, Application US/08083590A
: Patent No. 5786158
: GENERAL INFORMATION:
:   APPLICANT: Artavanis-Tsakonas, S. et al.
:   TITLE OF INVENTION: Therapeutic And Diagnostic Methods
:   TITLE OF INVENTION: And Compositions Based On No. 5786158ch Proteins And
:   TITLE OF INVENTION: Nucleic Acids
:   NUMBER OF SEQUENCES: 21
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Pennie & Edmonds
:     STREET: 1155 Avenue of the Americas
:     CITY: New York
:     STATE: New York
:     COUNTRY: U.S.A.
:     ZIP: 10036
:   COMPUTER READABLE FORM:
:     MEDIUM TYPE: Floppy disk
:     COMPUTER: IBM PC compatible
:     OPERATING SYSTEM: PC-DOS/MS-DOS
:     SOFTWARE: Patentln Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/08/083,590A
:     FILING DATE: 25-JUN-1993
:     CLASSIFICATION: 435
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Mistock, S. Leslie
:     REGISTRATION NUMBER: 18,872
:     REFERENCE/DOCKET NUMBER: 7326-015
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: 212 790-9090
:       TELEFAX: 212 8698864/9741
:       TELEX: 66141 PENNIE
:   INFORMATION FOR SEQ ID NO: 20:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 2556 amino acids
:       TYPE: amino acid
:       STRANDEDNESS: single
:       TOPOLOGY: unknown
:     MOLECULE TYPE: peptide
:     US-08-083-590A-20
:
Query Match      6.1%; Score 480.5; DB 1; Length 2556;
Best Local Similarity 21.7%; Pred. No. 2,3e-30;
Matches 269; Conservative 120; Mismatches 445; Indels 407; Gaps 75
:
0Y      127  EKVLEIQRKRCNNPTIITVRGEC--GKCSQQAACPPL-----TKPLRETR----- 169
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      256  ENIIDDCPGNCKNG-----GACVDSGVNRYNCPCEPMTGQYCTEDVDCQLMPNACQG 309
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
0Y      170  -----KCIYSIVFMGRSRVFIQGPDCVPTI-----ITRACWLASLAHNAK----- 210

```

```
Db 310 CTCNHTGCGYNCVYNGWGTGE-----DCSENIIDDCASACRHGATCHDRVASFYCE 360
QY 211 -PAPGEVAMCALGTASV-----WDGVNNGTGTCCQGLGFNGTAC---ETCTEGK 254
Db 361 CPHGRTGLLCHINDACISNPCNEGSNCDTNPYNKAICTCPGSGYGPACSQVDDECSLIG- 419
QY 255 YGICHDQACSCVHGRCSSQPLDGSDDCDVGRGKCDMEITTDNC-NCTGHTSANCLLD 313
Db 420 -ANPCEHAGKCTIN-----TLGSFECQCLQGYTGPRC--EIDVNECVSNPCONDATC-LD 469
QY 314 PDGKASCKCAAGFRNGTVCATAINA--CETS---NGGCTKADCKRTTPGNRVCCKAG 367
Db 470 QIGFEQCMCMPEYEG---VHCEVNTDECASSPCLHNGRIDKIN-----EFOECEPTG 519
QY 368 YTGDSIVC-LEINPCLENHNGGCDRAECTGTGPRNCAVNCNLPKRYTGDKVCSLINVCLTN 426
Db 520 FTGH--LCQYDDEBCAST--PCKNGAKCLD-GPNFTYTCVTEGYTG-----THCEVD 566
QY 427 NGGCSPPAFANCYTEODRI---CTCKPDYTGDIYCRGSIYELPKNPSTSOYFEOLE 482
Db 567 IDECDRDP-CHYSGCKDQAVFTGLCRPGYTGHH--CETNI--NEGSSQPCRLRGTCQDDP 622
QY 483 HA-----VRELAGPGLTFVAPRLSSFNHPRIKIDMDQGLMSQYLRYHVGCQQLLDN 537
Db 623 NAYLCFCLKGTGP-----NCEINLDD-----CASSPCDS 652
QY 538 LKVTTSATTLGCEPYSISVSQDTFVINEAKYLSDDISTNGVIIVIDKLSPKULLTP 597
Db 653 ---GCLDLKIDYECACEPRTGYTSMCNIDECAGNPCHNG------TC 693
QY 598 KDALGRVLQNTLTVAANHGYTKFSKLIODSGLSVITDSIHPTVTFWPTDKALELALPE 657
Db 694 EDGI-----NGFTGCPREGY-----HDPTCLSEVNECNSMPCV- 726
QY 658 QODPLFNODNKDKLSYLFHV-----IRDSKALASDLPSPASAKTLOGSELSVR 707
Db 727 ---HCACRDSLNGY-KCDDCPQMSGTINCIDINNECESNPVCVNGTCKDKMT-SGIYCT 778
QY 708 CGTG-----SDIGELFLINBQMRFIHRGLFDYGVAVYGDICLLMNPILGRCOTFTTF 760
Db 779 CREGSGNRCNQTINIECASNP--C--LNKGTCTIDVAGKCNCLL--PYTGATCEVLAP 832
QY 761 DIPGEC--GSC-----IFTPKCPILSKPKGVK-----KKCIYPLPFRNRVEGC 802
Db 833 CAPSPCRNKGEGROSEDEYESFCVP---TAGAKGQICEVDINCEVLSPC---RHGASC 885
QY 803 QNLCTVVIOTPRCHGHTMPDCQACPRGPD-----TPCNNRGMCRD-LYTPMGQC 851
Db 886 QN-----THGYRCHQOAGYSGRNCETDIDCRPNPCHNGSGCTDGINIAF--C 932
QY 852 LCHTGFNGTACE-----LCWHGRFG-----PPCOPRS 878
Db 933 DCLPGFRGTFCEDINIECASDPCRNAGNCTDCVDSYTCCTCPAGFSGICHENNPTDTESS 992
QY 879 CSEHQCDGEGITGSGECLCTETGWTAAASCDTPTAFAVC--TPASVHAHTEN---NTCVC 934
Db 993 CFNGSTCYVDGI-NSFTCLCPRGFTSGYQ---HVVNEEDSRCLLGLGTQDGRGLHRCIC 1048
QY 935 NLNTEGDDITC-TVYDFPKONNGGCAKAYAKSQKTQVSCSKKAYKG-----DGYSCIE 988
Db 1049 POGYGP--NQONLVHMC--DSSPCKNGGKCMQWOTIYRCECPSGMTGLYCDVPSVSC-E 1103
QY 989 IDPCADGVNGG--CHEHNTCMTGPKHKHCKECSHYVG---DGYD-CEPEDLPLDRCLQ 1041
Db 1104 VAAQRQGVAVARLCOHGLCYDAG-NTIHCRCQAGTGSYCDLVDDECS- 1152
QY 1042 DNGQCHPADSCADLYFODTTVGVFHLRSPGLQYKLTLPDKAKACAKAATIAITYNQLSYA 1101
Db 1153 --SPQNGATGTDY-----LGGYSCKCVAGYHGVNCSFEIDECLG- 1190
QY 1102 OKAKYHLCSSAGLWESGRAVYPTTYASQKCGANVGI-----VDGSRANKSEM 1149
```

```
Db 1191 -----HPCQNG---GTCLDLPNTYKC-SCPRTGQVHCEINVDNCPVPDYSRSPK--- 1238
QY 1150 WDVFCY-----RMKDVNCTCKAGYVSDGFSQGLDVL 1183
Db 1239 ---CFRNGTCVDQYGGYSCCTCPGVGE--RCEGVNECL 1273

RESULT 3
US-08-532-384-20
; Sequence 20, Application US/08532384
; Patent No. 6083904
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, S. et al.
; TITLE OF INVENTION: Therapeutic And Diagnostic Methods
; TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
; TITLE OF INVENTION: Nucleic Acids
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/532,384
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/083,590
; FILING DATE: 25-JUN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: MISTOCK, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7326-015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 8698864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2556 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-532-384-20

Query Match 6.1%; Score 480.5; DB 3; Length 2556;
Best Local Similarity 21.7%; Pred. No. 2.3e-30;
Matches 269; Conservative 120; Mismatches 445; Indels 407; Gaps 75;

QY 127 EKVLEIQRNCRNDTIYRGEC--GKSGQAPOPLE-----TYPLRETR----- 169
Db 256 ENIDDCPGNNCKNG-----GACVDDVNTYNCPCPEWTGYCTEDVDECQLMPNACQNG 309
QY 170 -----KCIYSIYFMGRKRSVFIQCQPOCVRII---ITRACMLASLAHNAK----- 210
Db 310 GTCNHTGCGYNCVYNGWGTGE-----DCSENIIDDCASACRHGATCHDRVASFYCE 360
QY 211 -PAPGEVAMCALGTASV-----WDGVNNGTGTCCQGLGFNGTAC---ETCTEGK 254
Db 361 CPHGRTGLLCHINDACISNPCNEGSNCDTNPYNKAICTCPGSGYGPACSQVDDECSLIG- 419
QY 255 YGICHDQACSCVHGRCSSQPLDGSDDCDVGRGKCDMEITTDNC-NCTGHTSANCLLD 313
Db 420 -ANPCEHAGKCTIN-----TLGSFECQCLQGYTGPRC--EIDVNECVSNPCONDATC-LD 469
```

Oy	314	PDGKCKCAAGRGVGTCTAIIIN	-CETS-----NGGCSFRKADCKRTTGNRVCVAKG	367
Db	470	QITEFQCMCPRETEG----	VHCEVNTIDECASSPCLHNGCLDKIN-----EFOCEPRTG	519
Oy	368	YTGDDGIVC-LEINPCLENHGGCDRNAECTOTGPNQAVCNCLPKYTGDBKVCSLINCLTN	426	
Db	520	FTGH--LCQYDVDECAST--	PKGNAKCID--GPNITYCTVCTEGYTG-----THEEVD	566
Oy	427	NGGCSPPAPCNTLEQORI----	CTCKPRYTGDIIVCGSITYGELPRKPNRSTSYFPLOE	462
Oy	483	HA-----VRELAPGPFVAFAPLASSPNEPRITKIMDOQGLMSQVLRYHVUGCOLLLDN	537	
Db	567	IDBCDDP-CHYGSCKDVAFTCTCLRPYTGHN--	CETNI--NECSSOPCLRGTCCDDP	622
Db	623	NAIYCLCKTTP-----	NCEIMDD-----CASSPDS	652
Oy	538	LKVYTSATTLQGBPVSISVSODTFINNEAKVLLSDIISTNGVIHVIDKLSIPKXLLITP	597	
Db	653	---GTCLDKIDGECACBEGYTGSMCNISIDECAGNPCHNG-----	TC	693
Oy	598	KDALGRLOULTVVAANHGYTKFSKLIQDSGLLSTITSIHPTVYVFWPTKALEALPE	657	
Db	694	EDGI-----NGFRCRPEG-----	HDPTCLSEVNECNSPCV-----	726
Oy	658	QODPLFMQDKDKLSYLFHV-----	IRDSKIASLDPRSASWKTLOGSELSVR	707
Db	727	-----HGACRDSLNG-KCDDCPGMSGTNCIDINNECESNPVCVAGKCKMT--	SGIVCT	778
Oy	708	CGTG-----	SDIGELFLNEOMCFRINHGLFIDVGAVALDCLLIMPTLTGRCDDTFPTF	760
Db	779	CREGFSGPCNQTINIECASN--	C--LNNGTICIDVAGIKCNCLL--PYTATEEVYLA	832
Oy	761	DIEPEC---GSC-----	IFPKCPCLSKPRGVK-----KKCIYNLPFRNREGC	802
Db	833	CAPSPCKNGEBCROSEDEYSEFSCVP----	TAGAKQICEVDINECVLSPC--RHGASC	885
Oy	803	QNLCTVVIOTPRCOHGYFMPDCACPGPD-----	TPCNNRGMCRD-LYTPMGQC	851
Db	886	QN-----	THGGYRCHCOAGSGRNCETDIDCPRNCHNGSGCTGINTAF--	C 932
Oy	852	LCHTGFNGTACE-----	LCNHGRG-----EDCPRS	878
Db	933	DCLRGFGTGCCEBDINECASDPRCANGACTDVSYTCTCPAGFSGICENNTPTRESS	992	
Oy	879	CSEHGOCDEGTSGSECLCTGTWTAASCDPTPAVAVC--	TPACSVHATCTEN--NTVCV	934
Db	993	CFNCGTCVDGI--NSFTCLCRPRTGTCQ--	HVNEDCSRPCLLGLGTCOORGRLHRTC	1048
Oy	935	NLANTEGDITC-TIVDFCKONNNGCAKVAKCSOKTGVSCSKKGYK--	DGYSIE	988
Db	1049	PGYTGPR--NQMLVHMC--	DSSPCKNGKCMQTHQYRCEPSCMTGIYCDVPSVSC-E	1103
Oy	989	IDPCADGVNGG--	CHENATQRMTPRKHNCSEKSHVUG-----DGYD-DEBDPLYDRLO	1041
Db	1104	VAAQROGVDAARLCQHGLCLVDAG--	NTNHCROAGTGTGSEDLVDECSP-----	1152
Oy	1042	DNGGCHPDASCADLYFODTTVGVFHLRSPJGQYKILFDFAKACAKKAAATATYNOJLSYA	1101	
Db	1153	---SPFCQNCATCTD-----	LGGYSCKVAGYNGVNCSEIDECLG-----	1190
Oy	1102	QKAYYHILCSAGWLESGHVAIPTTYASOKCGANYGI-----	VDGSRANKSEM	1149
Db	1191	-----HPCQNG--	GTCLDLPNTYKC--SCPRTQGHCEINVDNCPNPDVPSRSPK---	1238
Oy	1150	MDVFCY-----	RMKDVNCTCKAGYVCGFGSSGGLLOVL	1163
Db	1239	-----CFNNGTCVDVGYSCTCPRGVGE--	RCEGVNBNCL	1273

RESULT 4
US-08-185-432-19
; Sequence 19, Application US/08185432

```

? Patent No. 5750652
?
? GENERAL INFORMATION:
?
? APPLICANT: Artavanis-Tsakonas, Spyridon
? APPLICANT: Busseau, Isabelle
? APPLICANT: Diederich, Robert J.
? APPLICANT: Xu, Tian
? APPLICANT: Matsuno, Kenji
? TITLE OF INVENTION: DELTAE PROTEINS, NUCLEIC ACIDS, AND
? TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
? NUMBER OF SEQUENCES: 23
?
? CORRESPONDENCE ADDRESS:
? ADDRESSEE: PENNIE & EDMONDS
? STREET: 1155 Avenue of the Americas
? CITY: New York
? STATE: New York
? COUNTRY: U.S.A.
? ZIP: 10036-2711
?
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: PatentIn Release #1.0, Version #1.30
?
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US//08/185,432
? FILING DATE: 21-JAN-1994
? CLASSIFICATION: 530
?
? ATTORNEY/AGENT INFORMATION:
? NAME: Mistrock, S. Leslie
? REGISTRATION NUMBER: 18,872
? REFERENCE/DOCKET NUMBER: 7326-006
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: (212) 790-9090
? TELEFAX: (212) 869-8864/9741
?
? TELEX: 66141 PENNIE
?
? INFORMATION FOR SEQ ID NO: 19:
?
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2703 amino acids
? TYPE: amino acid
? TOPOLOGY: unknown
?
? MOLECULE TYPE: protein
?
? OS-08-185-432-19

```

[illegible]

Db 679 NE-----CHSNPC-----NNGAT-- 691
QY 571 SSDIITNGVIVHIDKLSPKLLI---TPKDALGRVLQNTTVAANHGYTKFSKLIDS 627
Db 692 -----CIDGINSYKQCVPGFTGQHCENKDECISSPCANNQ----- 728
QY 628 GLLSVTDSIHTPVTVFWPTDKALEALPPEQDLEFNQDNKDKLSYLKFIHVRDSCALE 687
Db 729 -----VCIDQV-----NGYKCECPRGFYDAHCLSDVDECA 758
QY 688 SDUPRSASMKTLQGLSELSTVRCGTG-----SDIGELFLBQMKRFTHRGLDVGAY 739
Db 759 SN-PCVNEGRCEDDINEFICHPRGYTGKRCCELLIDECSSNP--CQ--HGTCYDKLNAF 813
QY 740 GIDCLLMPFLTGRCRDPFTTFDIPGEC--GSCI-----FTPK-CPLKSPK 782
Db 814 SCQO--MRYTGQKCEITNIDCVINPCNGNGTCLDKVNGYKCVCKVPTGRCESKMDPC 871
QY 783 GVKR-----KCI--YNPLPR-----RNVEGQNL----- 805
Db 872 ARNRCKNEAKCTPSSNFDSCCTCKLGYTGRYCDDEDIDECSSLSPCRNGASCLANVPGSYR 931
QY 806 CTVIYIQRRCCHGYFMRDC-----QACRGGPTPCNNRGMCDLITPMG--QCLCHTNG 859
Db 932 CL-----CTKYEGRDCAINTDCC--ASFPCQNGRTCLD--GIDYSCICLVCDGFDG 978
QY 860 TACEL-----CMHGRF-----CPDCQPRSCSEHGQCD 886
Db 979 KHCTEDINECLSPQCNATQSOYVNSTYTCTPLGFSINGQINQINDEDCTESSCLNGGSCI 1038
QY 887 EGITGSECLCETGWTAAACDPTPAVAVG--TPACSVHATCTENN--TCVCNLYEGDG 942
Db 1039 DGINGY-NCISLAGYSGANCQYK--LNKCDNPLCATCHEQUNNEYTCHCPSFTGK- 1093
QY 943 ITCT-VVDFECONNGGCAKVAKCSQKQVSCSKCKYKGGYSC-IEIDCAGV----- 996
Db 1094 -QCEYVDWCGOS--PCENGATCSOMKHQFSCKSAGWTGK--LCDVYTISCDAAADKRG 1148
QY 997 -----NGGCHHATCRMTGPGKHKECKESHVYGDVDEPBOLPLDRLQDNQCHP 1048
Db 1149 LSLRQLCNNGTKDYG-----NSHYCYSOGYAGS--YQKE--IDEC--QSQPCON 1194
QY 1049 DASGADLYFDQTTGVPHLSPLQYKLTTPDKAEACAKAATATYANQLSYAQAKAYH. 1108
Db 1195 GGTGRDL-----IGAYEQCQROGQGNCELNIDDCANP----- 1229
QY 1109 CSAWLESGRAYPTTASOKGANYGVITVYGRANKSEMDVFCY-----RMKDVN 1161
Db 1230 CONGCTCHDRV---MNFSCSCPRTMGI--CEINNDCKPGACHNNGSCIDRVGGFE 1282
QY 1162 CTCRAGYVGDGFGSGNLLQYL 1183
Db 1283 CVCQPGFY--GARCEGDI NECL 1302

RESULT 5
US-08-185-432-17
; Sequence 17, Application US/08185432
; Patent No. 5750652
; GENERAL INFORMATION:
; APPLICANT: Artavanis-Tsakonas, Spyridon
; APPLICANT: Buseau, Isabelle
; APPLICANT: Diederich, Robert J.
; APPLICANT: Xu, Tian
; APPLICANT: Matsuno, Kenji
; TITLE OF INVENTION: DETEX PROTEINS, NUCLEIC ACIDS, AND
; TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York

COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 2556 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-17

Query Match 6.0%; Score 471.5; DB 1; Length 2556;
Best Local Similarity 20.6%; Pred. No. 1.3e-29;
Matches 262; Conservative 113; Mismatches 427; Indels 471; Gaps 73;

QY 127 EKVLEIQKRNQDNNDTIIVRGC--GKCSQAPCLE-----TPKLAETR----- 169
Db 256 ENIDDCPPNNCKNG-----GACVDGVNTYKPCPEPTGQYCTEDVDECOIMPACONG 309
QY 170 -----KQIYSIYEMGRKSVFTGQPCQCVRTI--ITRACWLASIAHNAK----- 210
Db 310 GYCHNTFHGGVNCVCMNGTGE-----DCSENIIDCASAACHGATCHDRVASFCE 360
QY 211 -PARGEVKCALGTASV-----WDGVNGTGTQCGIGLFGNGTAC---ETCTEGK 254
Db 361 CPHGRTGLLCHLNDACISNPNEGSNCDTNPNGKAICTCPSGYNGPACSDVDDECSLG- 419
QY 255 YGIHDDQACSVHNGSCGRLDGSQCDQDVGRGVKCMELITTMNC-NGTCHTANCLLD 313
Db 420 -ANPCEHAGKCTN-----TLGSEFCQCLQGYTGPRC--EIDVNCVSNPCQNDATC-LD 469
QY 314 PDGKASCCAGAFRNGNTVCTAINA--CETS-----NGCSTKADCKRTTPGHRVCCKAG 367
Db 470 QIGEQCMCMRGYES---VHCEVNTDEBCASSPLCHNGKLDKIN-----EFGCECTG 519
QY 368 YTGDIIVC-LEINPCLNHHGCDRNAECTGTGPNQAVNCLEPKYTGDKKVGSLINVLITN 426
Db 520 FTGH--LCQYDVDECAST--PCKNGAKCLD-GPNITYTCVCTEGYTG-----THCEVD 566
QY 427 NGGCSFPAFCWYTEDQDI-----CTCKRPTYG-----DGLVCGSIYELPKRPS 472
Db 567 IDECDPDP-CHYGCKDQVANFTCLCRPGYGHHCETNINECSSQPCR-LMGTC-QDPD 622
QY 473 TSOYFFQLQEHAVRELAGPGFTVAFAPLSSSFNHEPRITKMDQQLMSQVLRHYHVGQO 532
Db 623 NAYLCE-----CLKCTGP-----NCEINLDD-----CAS 647
QY 533 LLDNLKVTTSATTLQGPVSIYSQDTVFJNNEAKVLSDOIISTNGVIVHIDKLSPKN 592
Db 648 SPCDS--GTCLDKIDGYEACCEPGYTGSMNSNIDECAGNPHNGG----- 691
QY 593 LLIIPKDALGRVLQNTTVAANHGYTKFSKLIDSGLLSVTDSIHTPVTVFWPTDKALE 652
Db 692 ---NCEDEGI-----NGFYCRPEGY-----HDPYTCLEVNECSNCPV----- 726
QY 653 ALPPEQDLEFNQDNKDKLSYLKFIHVRDSCALESDLPRSASMKTLGSGELSVRC----- 708

Db 727 -----HGACWMSLNG-----YCCDDP 743
Qy 709 ---GTGSDIGELFLNEQMC-----RFTHR 729
Db 744 GWSGNCINNNECSBNSGNGTCKDMTSGIVCTWEGFSGPNCOTINIENCSNPCLK 803
Qy 730 GLEPDVAVAGIDCLLMPRTLGRCDDTFTFDIPGEC---SSC-----JFTPRKPLKS 779
Db 804 GTCIDVAVAGYCNCLL---PYTGATCEVVLAPCAPSPCRNGEGECROSEDEYESTSCVP--- 858
Qy 780 KPKGVK-----KKCIYNPLPFRNVEGCONLCTVYIOTPRCHGYFMPDQCAPGP 831
Db 859 -TAGAKGQTCEDINECVLSPC---WHGASQCN-----THGXRYRCHQAGYSGR 903
Qy 832 D-----TPCNRBMCRD-LYTPMGOCCLHTGFNGTACE----- 863
Db 904 NCFDIDDCWPNPCNNGSCTDGINJAF---CDCLPFGWGTCEEDINECASDPCCRNGANC 961
Qy 864 -----LCMHGRFG-----PDCQPRSCSEHGOCDEGITSGECLCETGWTAAACD 907
Db 962 TDCVDSYTCCTCPAGFSGIHCENNTPPDCTESSCFNGTGYDGI-NSFTCLCPGFTGISTCQ 1020
Qy 908 TPTAVFAVC-TPACSVHATCTEN---NTVCNLYNEGSDGITC-TVVDFCKQNGGCAVVA 962
Db 1021 ---HVNVNCDRPPCLLGGTCQDGRGLHRCCTPOGYTGP---NQCNLYVMHC---DSSPCXKNG 1073
Qy 963 KCSQKGTQVSSCKKGYG-----DGYSCIEFDPCADVNGG---CHEHATCMTPGKXK 1015
Db 1074 KCMQTHQYRCBPBGWGLYCDVPSVC-EVAAQROGVDAVARLQHGGLCVDA-CNTNH 1131
Qy 1016 CEKSHYVG---DQVD-CEPEOLPLDRCLODNGQCHPASCADLYFODTIVGVFHLSP 1070
Db 1132 CRCAGYGVSGEDLYDCSP-----SPCQNGATCTD-----LGGSCKCV 1173
Qy 1071 LGQYKLTFDKAKKACAKAATATATYINQLSYAKAKYHLCAGWLESQVAPPTTYASQK 1130
Db 1174 AGYHGVNCSSEIDCLSL-----HPCQNG---GTCLDLPNTV---KC 1208
Qy 1131 GANVGIYDGSBRANKSEM-----MDVFCY-----RMADNCTGACVYG 1170
Db 1209 SC-----PMGTQVHCINVDCCNPVDPVSWSPKCFNNGTCVDQVGGYSGTCLPPGFGV 1262
Qy 1171 DGFSCSGLLOVL 1183
Db 1263 E--RCEGDVNECL 1273

RESULT 6
US-08-185-432-16
Sequence 16, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
NUMBER OF SEQUENCES: 23
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
CORRESPONDENCE ADDRESS:
ADDRESS: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-16

Query Match 5.7%; Score 450.5; DB 1; Length 2471;
Best Local Similarity 22.7%; Pred. No. 6.9e-28;
Matches 270; Conservative 108; Mismatches 421; Indels 391; Gaps 73;

Qy 133 OKNRCDNDITIIIVREGCKC-----SQQAPCLETKKPL-----RET 168
Db 69 EKNRQNGCTCYAQMALKRATCRASGFTGEDCOYSTHPCFVSPPCLNGTCHMLSDT 128
Qy 169 KCIYSIYFNGKRSYFICQPOCVRTIITRACMLASLAHNAKPAPGEVYKCALGASVMD 228
Db 129 YECQVQVFTGKE-----CQ-----WTDAC-----LSH-----PCANG---STCT 160
Qy 229 GYNGTGCQCGGFGNGTACET-----CTGKGYIHDDO 261
Db 161 TVANQFSCKLTGFTGQCEFDVNECDIPGHQHGCTCLNLPSTYQCCPGGFTGYQYDS 220
Qy 262 ---ACS---CVH-GRCSGGLPDDGS---CDDVGNRGVRCMEITTDNC-NGTCHTSANCL 311
Db 221 LVPAPAPSPCVNNGGICRQ---TGDFTEECNCLPGEFGSTCERNI---DDCPNHRQNGVGYCV 276
Qy 312 LDPDGKASCKCAAGFRNGVYCTA--INACETSNNGCSTKADCKRTTPGNRYVCYKAGYTG 370
Db 277 ---DGVNTYNGRCRPPQWTCQFCTEDVDECLLDPPNACQNGGTCANRN--GGYGCVCYVNGSG 332
Qy 371 DGIYCLEINPCLENHGGCDRNAECTQGTG---PNOAVNCCLPKYTGSGKVSILNVCLNN 427
Db 333 D-----DSENIIDC-ARASCTPGSTCIDRVASFSCMCEGAKAGLCHLDACISN- 382
Qy 428 GGCSPFAFCNYTE--ODQRICTCKPDYTG-----DGIYKRSIYGLPKNPSTSQYFFQ 479
Db 383 -PCHKGALCDINPLNGYVICTCPQGYKADCTEDVDECBAMNS-----NPC----- 427
Qy 480 LQENAVRELAPGPPTVAPLSSSFNHEPRIKDWQDQIMSOVLRYHVVGCOQ---LLLDN 537
Db 428 --EHAQKCVNTDGAFA-----HCECLKGYAGRPCENDINCHSDPQONATATCDK 474
Qy 538 LKVTTSATT--LOGRPVSISSV--ODYVFINNEAKVYLSSDIS-----TNGVYHV 583
Db 475 IGGFTCLCMPKFGVHCELEINECOSNPCVNNQCV---DAVNRFOCLCPGPGFTGPVQOI 531
Qy 584 -IDKLSPKNLLITPKDALGRVLONLTVVAAHNGY-----TKFSKLIQDSGLSVITDSI 637
Db 532 DIDCCS-----TP-----CLNGAKKIDHPNNGIECCATGFTGVLCENINDNDPPPC 579
Qy 638 HTPVTVFWPTDKALEALPPEODPLFNQDNKDKLSYLKFIYIRDSKALASDLPASWK 697
Db 580 H-----HGOCODGIDSYCTICINPGYMAICSD----- 606
Qy 698 TLQGSSELSVRGCTGSDIGELF-----LNQOMKRFTHRGLFPDVGAVAYGIDCLLMNPTLGR 753
Db 607 -----QIDECYSSPCLNDGRC-----IDLNVGYQCNQCPGPTSGVN 641

Db 843 ENAAVCKESPNESTYTCACAPGMOGO--RCTIDIDEC---ISKPCMNHGLCHNT--QGSYM 896
QY 1016 CECKSHYVGDVDCPEPOLPLDRCLODNGQCHPDASCADLYQDPTVGFHLRSPGLQYK 1075
Db 897 CECPPGPF--SGMDCEED---IDDCLAN--PCONGSGCMD-----GVNPF---SCLCLPG 940
QY 1076 LTFDKAK-----EACAKEAATATYNOJLSYAOKAKYHLCAGWLESGRVAYPTIYAS 1127
Db 941 FTGDKCQIDMNECLSEPC-KNGGTCSDYVN-SYTKC-----COAG-----FDG 981
QY 1128 QKCGANVGIYDYSRANKSEMMDVFCYR-----MKDVNCTCKAGYVG 1170
Db 982 VHCENNI-----NECTESSCFNGGTCVGDGINSFSCLCVPGFTG 1019

RESULT 8
US-08-532-384-19
Sequence 19, Application US/08532384
Patent No. 6083904
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, S. et al.
TITLE OF INVENTION: Therapeutic And Diagnostic Methods
TITLE OF INVENTION: And Compositions Based On No. 6083904ch Proteins And
TITLE OF INVENTION: Nucleic Acids
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/532,384
FILING DATE:
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/083,590
FILING DATE: 25-JUN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-015
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 8698864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2471 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-532-384-19

Query Match 5.7%; Score 450.5; DB 3; Length 2471;
Best Local Similarity 22.7%; Pred. No. 6.9e-28;
Matches 270; Conservative 108; Mismatches 421; Indels 391; Gaps 73;

QY 133 QKNRCDNDNTIIVRGECKC-----SQAPCELEKPKL-----RET 168
Db 69 ENNRONGSTCAVQAMLKATCRCAAGTGEDCOYSTSHPCFVSRCCLNGTCHMLSRGT 128
QY 169 KCIYSIYFMGRSVFIGCQPOCVRTIITRACWLASLAHNAKPADGEVWKALGTAASVMD 228

Db 129 YECTQVGFSGNE-----CQ-----WTDAC-----LSH-----PCANG--STCT 160
QY 229 GVNGTGTCCQCGFNGTACET-----CTESKYIHCQ 261
Db 161 TVANPFSCKCLGTGFKCEETVNECDIPGHQHGSTGLNLPGSYQCCQPOGFTQGYDS 220
QY 262 ---ACS---CVH--GRCSGGLDGS--CDQDVGMRGVKCDMEITTDNC--NGTCHTANCL 311
Db 221 LTVPCAPSPCVNGVGTGRQ--TGDFTECNCLPGEFSGTCEINI--DDCFNHRQMGVCV 276
QY 312 LDPDGKASKCAAGFRGNCTVCTA--INACETNSGGCSTKADCKRTTPGRVRVCVACAGYTG 370
Db 277 --DGVNITVNCPCPQWMTGQFCTEDVDECLQPNACQNGTCAANN--GGYGCVCVNGSG 332
QY 371 DGIYCLEINPCLENHGGCDRNAECTQTG--PNAVCNCLPYRTDGGKCSLINCLTN 427
Db 333 D-----DCESENIDDC--AFASCTPGSTCIDRVASFSCMEPEKRAGLLCHLDACISN- 382
QY 428 GGCSPFAPCNYTE--ODRITCTCKPDYTG-----DGIYCRGSIYELPKNPSTQYFQ 479
Db 383 -PCHKGALCDTNPNGOYICTCPQGYKGGADCTEDYDECAMANS-----NFC----- 427
QY 480 LOEHAVRELAGPPTVTFAPRLSSFNHPRIKMDQQGLMSQVLRHYVGGQ--LLLDN 537
Db 428 --EHAQKCVNTDCAF-----HCECLKYVAGPCEMDINCHSDPCONDAATCLDK 474
QY 538 LKVTTSATY--LOGEPVSIYS--ODTVEINNEAKVLSIDIIS-----TNGVIHV 583
Db 475 ICGFTCLCMRPGKGVHCELINEQSNPCVNNQCV--DKVNRQCLCPGFTGVCQI 531
QY 584 -IDKLISPKNLLITPRKDALGRVLQNTTYAANHGY-----TFESKLIJDSGLSITBISI 637
Db 532 DIDDCSS-----TP-----CLNGAKCIDHPNGECCQCATGTGVLCEENIDNDPPDC 579
QY 638 HTPVTVFMTKALEALPEEQDQDFLENDNKKLSYLKFHVIIRSKALASDPRSASWK 697
Db 580 H-----HGOCQGDISTYTCINPGTGAICSD----- 606
QY 698 TLQSELSVRCGTGSDIGELF--LNEQMCRFIHRGLLFDVAVYAGIDCLLMNPLTGR 753
Db 607 -----QIDBCYSSPLNDGRC-----IDLNVGYCNCQPGTSGVN 641
QY 754 CDTFTTFD-----IPGEC-----GSCITFP-----KPLKSKPKGVKKKCIYNPLPF 795
Db 642 CE--INFDCASNPCIHGICMDGINRYSCVSPGFTGQCNIDI-----DDCASN-- 690
QY 796 RNRVEGQNLCTVVLQTPRCHGYFMPDQOAPGGGDFPCNRMGRCDLYTPMGQL-- 852
Db 691 -----CRKGAT-----CLNGVNGFC--TCPEGFHP-----SCYSOVNECLSNP 728
QY 853 -----CHTGFNGTAGELCMHGRFRGPDQ-----PRSCSEHQQCDDEGITGSGECLCTG 900
Db 729 CLHGNTGSLGSYKC--LQAGWVGINCEYDKNECLSNPCQNGSTGDNLVNGY--RCTCKRG 786
QY 901 WTAASCDTPTAVFAYCTPRCSVHAHTEN--NTCYCNLNTYAGDGTTC--TYVDFEKQNNG 956
Db 787 EKGYNQ--VNIDECASNCLNQGTCFDDISGYTCHVLPYTK--NCQTVLAPSPBNPC 842
QY 957 GCATYAKCSQKGTQVSCSKKGYKGDGYSC--LEIDPCADGVNGGCHENATCMTGSGKIR 1015
Db 843 ENAAVCKESPNESTYTCACAPGMOGO--RCTIDIDEC---ISKPCMNHGLCHNT--QGSYM 896
QY 1016 CECKSHYVGDVDCPEPOLPLDRCLODNGQCHPDASCADLYQDPTVGFHLRSPGLQYK 1075
Db 897 CECPPGPF--SGMDCEED---IDDCLAN--PCONGSGCMD-----GVNPF---SCLCLPG 940
QY 1076 LTFDKAK-----EACAKEAATATYNOJLSYAOKAKYHLCAGWLESGRVAYPTIYAS 1127
Db 941 FTGDKCQIDMNECLSEPC-KNGGTCSDYVN-SYTKC-----COAG-----FDG 981
QY 1128 QKCGANVGIYDYSRANKSEMMDVFCYR-----MKDVNCTCKAGYVG 1170
Db 982 VHCENNI-----NECTESSCFNGGTCVGDGINSFSCLCVPGFTG 1019


```

1      RESULT          9
2      US-08-400-159-2
3      ; Sequence 2, Application US/08400159
4      ; Patent No. 5869282
5      ; GENERAL INFORMATION:
6      ; APPLICANT: Ish-Horowitz, David
7      ; APPLICANT: Henrique, Domingos M.P.
8      ; APPLICANT: Lewis, Julian H.
9      ; APPLICANT: Myal, Anna M.
10     ; APPLICANT: Fleming, Robert J.
11     ; APPLICANT: Artavanis-Tsakonas, Spyridon
12     ; APPLICANT: Mann, Robert S.
13     ; APPLICANT: Gray, Grace E.
14     ; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
15     ; TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON
16     ; NUMBER OF SEQUENCES: 20
17     ; CORRESPONDENCE ADDRESS:
18     ; ADDRESS: Pennie & Edmonds
19     ; STREET: 1155 Avenue of the Americas
20     ; CITY: New York
21     ; STATE: New York
22     ; COUNTRY: USA
23     ; ZIP: 10036-2711
24     ; COMPUTER READABLE FORM:
25     ; MEDIUM TYPE: Floppy disk
26     ; COMPUTER: IBM PC compatible
27     ; OPERATING SYSTEM: PC-DOS/MS-DOS
28     ; SOFTWARE: PatentIn Release #1.0, Version #1.30
29     ; CURRENT APPLICATION DATA:
30     ; APPLICATION NUMBER: US/08/400.159
31     ; FILING DATE: 07-MAR-1995
32     ; CLASSIFICATION: 435
33     ; ATTORNEY/AGENT INFORMATION:
34     ; NAME: Mirock, S. Leslie
35     ; REGISTRATION NUMBER: 18,872
36     ; REFERENCE/DOCKET NUMBER: 7326-029
37     ; TELECOMMUNICATION INFORMATION:
38     ; TELEPHONE: (212) 790-9090
39     ; TELEFAX: (212) 869-9741/8864
40     ; TELETYPE: 66141 PENNIE
41     ; INFORMATION FOR SEQ ID NO: 2:
42     ; SEQUENCE CHARACTERISTICS:
43     ; LENGTH: 1404 amino acids
44     ; TYPE: amino acid
45     ; TOPOLOGY: linear
46     ; MOLECULE TYPE: protein
47     ; US-08-400-159-2

```

```

Query Match          5.3%; Score 413.5; DB 2; Length 1404;
Best Local Similarity 19.7%; Pred. No. 3.4e-25;
Matches 274; Conservative 127; Mismatches 482; Indels 507; Gaps 66;

QY 237 OCGLGFNGACET-----CTEKGYGTHCDOA-----CSCVHGRC 270
      || : : : | | | | | : : : : : | : : : : |
Db 236 OCAYIYYINTTCTTTPRPDPDGFHAYACSGEGOKLLNGQVNCBEALICKAGCDPVHGRC 295
      || : : : : : | : : : : : | : : : : : | : : : : |
QY 271 SGGPLGDSGCDSDYMGWGVKCMDEITTTDNC-NGTCHTSANCLLDDPKASCKCAAGFRGN 329
      || : : : ||| : | : : : : | : : : : : | : : : : |
Db 296 DR-----PGCECECRPGMRBGLCNECWVYEGCKRIGSGNSA-----WMCVCEDTINNG 340
      || : : : : : | : : : : : | : : : : : | : : : : |
QY 330 GTVCTA-INACTSNMGCGSTKADCKRTTPGNFVYCYCKAKGYTGDGLVCLIEINPLENHGGC 388
      || : : : : : | : : : : : | : : : : : | : : : : |
Db 341 GILCQDLDLNFECT-HEPCKHGCTCENTAPDKYRCYCAEGLSGEOEIVE-HPCATR-PC 396
      || : : : : : | : : : : : | : : : : : | : : : : |
QY 389 DENACTQTGPNQA-----
      || : : : : : | : : : : : | : : : : : | : : : : |
Db 397 RKGIGCTLKTSNRTOAQYRTSHGGSNMCRPVRBSSMSRLDHLRPEGOALNGSSSGILV 456
      || : : : : : | : : : : : | : : : : : | : : : : |
QY 403 -----VCNCLPRYTGDKGVCSLLINCLITNNGGCSPPAFACNYTEODORICT 447
      || : : : : : | : : : : : | : : : : : | : : : : |

```

Dd	457	SLGSLQLOQOLAPDFTCDCAAGT--GPICE--INIDECAGGCEHGCTIDLIGFR--CE	512
Qy	448	CKPYTDGIVGRCISYIGELPKNPSTSOYFPFOJQEHAVELAGPGPFTVFAPLSSSFNHE	507
Dd	513	CPPEMHOD--VCOVD-----VNECAP-----	532
Qy	508	PRKIDWDOGLMSOVLRYHVWGCOOLLNKLKVTTSATTLOGEPVYSISODIYFINNEA	567
Dd	533	-----HSAG-----IAANALLTTPATVALISNLS-----STALLAALT	565
Qy	568	KVLSSDIITNGYIHVIDLKLLSPKNNLITPKNALGV--LOMLTIVA--AAHGTTKPSKL	623
Dd	566	SAVASTSLAIGPCINAKECRNQGSPACJCKEWMGVTCAEMLDDCQVQCRGATCID--L	624
Qy	624	10D-----SGLLSVITDSIHTPEVTJFEMPTDAXLEALPEQODPLFNQDNKDKLSYLF	677
Dd	625	VNDYRCACASGFTG-----	638
Qy	678	HVIDSKALASDLPRASAMKTLQGSLSYRCGTGSDIGELFLNMQCRFIRHGILLFDYV	737
Dd	629	---RDCE---TDIDECATSPCRNGE-----C-----VDM	662
Qy	738	AVGIDCLLMPNLTLAGRCDITFTFDIGEC--GSCITPR-----CPLSKRKGVKKKITY	790
Dd	663	VGKFNICICPLGYSGLCEAEKENCSTPCLGHCILNTPBGYYCHCP---PDRAGKC--	716
Qy	791	NPLFRNRNVEGCONLCTVVIQPRCHGF--	830
Dd	717	-----DOLRPLS---OPPCBEGCFANYSIATSAITTTTTTTTTTTATTTTRKMAKPSG	763
Qy	831	PDPFCNNRGMCR--DIATTPMGQCLCTHTGNGTACELCWMGRGPRPCOPRSCSEHGOCDEG	888
Dd	764	---LPCSHGSGEMSDVGT---PCKCHVGHGTFC---HNL--NDCSPNCRNGICLDG	813
Qy	889	ITGSGECLCEFGWTAASCPITPAVEAVCPACSAHATCTENNTCYCNLNEBGGITCTIV	948
Dd	814	DGDFTECMSSGWTG-----KCESERATGYCAQC-----	842
Qy	949	DPCRONNGGCAKAYAKSCOKGTQVSCSKRGYGDGSCIE--IDPCADGVNGCHEHATCR	1007
Dd	843	---QNGCTCMPGA--PDKALOPHCRCARGW--TGIFCEALIDCR---GQPCNHGTCIE	891
Qy	1008	MTGPKHRCCECKSHYVGDYDCEPEOPLDRLQDNQGCHPDASCADLYFDQTTGVFHL	1067
Dd	892	SGAGWRCVCAQGF--SGPDC--RINNECSPO--PCGGATCID-----GICGVC	937
Qy	1068	RSPLOGY---KLTPEKAEKAEKAEATATYNNOLUSTAOAKANYHLCASGWL--SGRYA	1120
Dd	938	ICPGRHGLRCIELLSDPKSACONASNTISPYALTNRSG-----NMLDIALTGR--	986
Qy	1121	YPTTYAOKCGANVGIYDYSRANKSEMDVFCYMKRYNCTCAAGYVGQFSGSGULL	1180
Dd	987	---TEDENCNA---CVCENGTSRCTNLW-----CGLPNCYKVPPLSSSNLS	1028
Qy	1181	QVLMSPSLTNFLTEYIAFAFSKASQAOAFILKHLTDLSIRGTLFVPONSGLPCKNSL	1236
Dd	1029	GVCKQHEVCYPAISETCLSSPCNWRDC--RALEPSRRVAPRILDAKSSCMPNOAVNEN	1086
Qy	1237	-----SGRDIEHLLTNVNSFYNDLVN--GTF-----LRTMLG	1267
Dd	1087	CARLLITLALERBGKASVBCGLCSILVRVLLAOLIKKAPASTGQDPGMVLVCLDKTCTN	1146
Qy	1268	SOLLITSSODLHO-----ETRPVDG--SLTIQNDIILAANGILHITSEPL	1310
Dd	1147	DIVELTVSSSKLNDPQLPVAVGLLGELSSRQNGQRKKELEIIOHAKLAALTSLIVEYKL	1206
Qy	1311	RAPPTAATAAHSIGTGFICAV--VVTGAIALAAYSYRLKORTT-----	1354
Dd	1207	ETAKVAOSGHS--LLIGVLCGVITLVGVSVFSLYWKRLAKYRISSGNLTPSLDALRH	1255
Qy	1355	-----GFQRPDKRTLMSMLAASSPRISOT--LCMPORRHQSPVPTPSQTLNRI	1405
Dd	1266	EEKSSNNLQNEENLRRTYTNPLKGTSSSLRATQMETSLNP-----APILAAASAASSATL	1319

QY 1406 WRTATLMGHC 1415
 1 : 1 :
 Db 1320 HRSQPLFPCC 1329

RESULT 10
 US-08-611-729A-2

Sequence 2, Application US/08611729A
 Patent No. 6004924

GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David
 APPLICANT: Henriquez, Domingos M.P.
 APPLICANT: Lewis, Julian H.
 APPLICANT: Myatt, Anna M.
 APPLICANT: Fleming, Robert J.
 APPLICANT: Artavanis-Tsakonas, Spyridon
 APPLICANT: Mann, Robert S.
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE
 TITLE OF INVENTION: SERATE GENE AND METHODS BASED THEREON
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/611,729A
 FILING DATE: 06-MAR-1996
 CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie
 REGISTRATION NUMBER: 18,872
 REFERENCE/DOCKET NUMBER: 7326-037
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 790-9090
 TELEFAX: (212) 869-9741/8864

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1404 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-611-729A-2

Query Match 5.3%; Score 413.5; DB 3; Length 1404;
 Best Local Similarity 19.7%; Pred. No. 3.4e-25;

Matches 274; Conservative 127; Mismatches 482; Indels 507; Gaps 66;

QY 237 OCGLGNGTACET-----CTEGYGTGHCDA-----CSCYHGRC 270
 11 : : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 236 QCAVYTYNTCTTCFRRDDDFGHVACGSEGOKLCLNGMGVCEBAICAKGCDPVHGRC 295
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 271 SGGPLGDSGCDGVGRGVKCDMEITDNC-NGTCHTSANCLLDPGKASCKACAFGRG 329
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 296 DR-----PGCEBEPGRGRLNCEMAYPGCKHSGCNSA-----WKCVCCTNMG 340
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 330 GYVCTA-INACETSNCGSTKADCKRTTPGNRVYCVCAKAGYGDGIVCLINPCLNHHGCG 388
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 341 GILCDLDFNFCGT-HEPCKHGTCENTAPDKYRCTCAEGLSGGCEIVE-HPCATR--PC 396
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 389 DRNAECTGTGPNDA-----402

Db 397 RKGCTCTLKTSNRTOAQYRTSHGRSNMGRPYRBRSSSMKSLDLHRLPEQALNGSSSGLY 456
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 403 -----VCNCLPKRYTGDKVCSLINVCNLNNGCSPFAPCNYEDQRICT 447
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 457 SLGSLQLOOQALPDPFTCDCAAGMT--GPICE--INIDECAGCGEHCGRGICDLIGGFR--CE 512
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 448 CKPDYTGDCIVCRGSLYGLPKNPSTSOYFFPOLQEHAVELAGPGRFYFAFLSSSFNIE 507
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 513 CPPEWHGD--VQAVD-----VNECBAP-----532
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 508 PRKIDMDQGLMSQVLRVHVCCOOLLNLTATTSATLQGEPIYSIVSDTFEINNEA 567
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 533 -----HSAG-----IAANALTTATATIIISNLS-----STALLALT 565
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 568 KYLSDDIISTNGVIVHYDKLSPKMLITPKDALGRV--LQNLITVA--ANHGRTFSKL 623
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 566 SAVASTSLAGPICINAKKECRNPGSPFACICKEGMGVTCANMLDPCVGCRRGATCID-L 624
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 624 IOD-----SGLLSYITDSIHPVIVFWPTDALEALPEQODLFENDNKKLSYLKE 677
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 625 VNDYRCACASGFTG-----638
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 678 HYIRDSKALASDLPKASMKTLQGSLSVRCGSDIGELFLNEQWCRFIHRGLLEFVGV 737
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 639 ---RDE--TDIDCATSPCKRNGE-----C-----VDM 662
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 738 AVGIDCLMNPPLGRCDFTFPDIPEGC--GSCITFPK-----CPKSPKGVKKCIY 790
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 663 VGFKNICPLGYSGLCEBAKEKNCPTSPCLBCHCLMTPGGYCHCP--PBRAGHC-- 716
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 791 NPLPRRNVGCGNICTVVIQTPRCCHGF-----MDCQACPRG 830
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 717 -----EOLRPLCS--QPRCEGCFANVSLATSATTTTTTTATTTTRKMAKPSG 763
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 831 PDPCCNRMGR--DLTPMGOLCHTGFNGTACELCWMGRGPRDQPPSCSEHGQCDG 888
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 764 --LPCSGHSCSEMSDGT--FCCKIVGHTGTFCE--HNL--NECSPPRCNGGICLDG 813
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 889 ITGSGECLCETGWTAAACDTPAVFAVCTPACSVHATCTENNTVCNLYEGDGLTCTV 948
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 814 -DGDFTCEGMSWGTG-----KRCSEKATGCVAGC-----842
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 949 DPCKQNNGCACAVAKCSOKGTGVSCSKKGYKGDYSCIE-IDPCADGVNNGCHEHATCR 1007
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 843 ---QNGGTGCMGA--PDKALQPHRCAPGW--TGELFEALIDOCR--GDPCHNGTGE 891
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1008 MTGPGKHCKECKSHYVGDVDEPEQLPLDRCLQDNGGCHPAPASCDLVLEQDTVGVFHL 1067
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 892 -SGAGWFRVCVCAQGF--SGPDC--RINVNECSPO--PCGGATCID-----GIGGYSC 937
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1068 RSPPLQY---KLTPEKAECAKEAATITATYNOLSYAQKAKYHLCSAGMLE--SGRYA 1120
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 938 ICPGRHGLRCELLSDPRASQOMASNTISPYTALNRSG-----NMLDIALTLGR-- 986
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1121 YPTTASQCGANVGIIVDYGSHANKSEMMDVFCYRMKDVNCTCKAGYVGDFSCGNLL 1180
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 987 ---TEDDEMCNA---CVCENGTSRCTNLM-----GGLPNCYVVDPLSKSNLS 1028
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1181 QVLMSPSLJNLFLEVLAVRSKSSARGQAFKLKHLTDLSTIGTLEVPNDSGLPGKSL---- 1236
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 1029 GYCKQHEVCVPALSETCLSSPCNVKGDG--RALEBRRAVAPRLPAKSSCMFNQAVNEN 1086
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1237 -----SGRDIEHLITNVNVSFYNDLVN--GTF-----LRTMLG 1267
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 1087 CARLITLLALERKGAASVGLCSLVRVLLAQLIKKPASTGQDQGMVLCDLKTGTN 1146
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1268 SOLITFSODOLHQ-----ETRYVDG--RSIIQMDIIAANGIIHITSBL 1310
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 1147 DIVELTVSSSKLNDPOLPAVAVGLAGELSSROLNGIQRKKELELOHAKLAALTSIVEKL 1206
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 QY 1311 RAPTATATAHSGLTGIGICAV--VLVTGAIALAASVYRLKORT-----1354
 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 : 1 :
 Db 1207 ETARVADSGSHS-LIIGVLCGVFIYLVGFSVFLISLYWKORLAYRTSSGMNLPISDLALRH 1265

Db 489 NLCQLDIDYCEBNPCONGAOCYNRASDYFCCKPREDEGNCNCHLKH-----CR----- 537
QY 540 VTTSATTLTGGEBVSISVODTFVFINNEAKVLSSDIITNGVTHVIDKLSPNLLITPKD 599
Db 538 -TTPEVIDISCTVAAM-SNDT-----PEGVATISSNVCGR----- 570
QY 600 ALGRVLQNLTTVAANHGYTKFKSLIODSGLSVITDSIHTPTVTPMTDKALEALPREQO 659
Db 571 -----HG----- 572
QY 660 DFLFNQDNKDKLSYKLFHVITRDSKALASDLPRASAMKTLQSGSELSVRCTGSDIGELFL 719
Db 573 -----KCKSQSGKFTCCDNKG-----FT 591
QY 720 NEOMCRFIHGRGLFDYGAAYGIDCLLMPNLTGRCDTFTTFDIPBGCSGCIPTPKPLKS 779
Db 592 G-----TYCHENI-----NDC-ESNPCRNG-----GTCT----- 614
QY 780 KPRGVKK-KCIYNPLFRFRNVEGCONLTCTVVIOTPRCHGYMPDCA-----CPGGPDP 834
Db 615 --DGVNSYKCI-----CSDGMEGAYCETNINDC-----SQNP 644
QY 835 CNNRGMCRDLYTPMGQCLCHTGFNGTACELCMHGRGPRDQPRSCSEHGQC-DEG1TGS 893
Db 645 CHNGGTCRDLVNDP-YCDCKNGMKKCTC---HSR-DSQCDEATCNGGTCYDEG--DAF 696
QY 894 ECLCETGWTAAACDPTTAVFAVCTP-ACSVNATCTENN---TCVCNLNTEGSGITCTVVD 949
Db 697 KCMCGMGWGTTCN--IANNSSCLPRCHNGGTGVNNGSFTCYCKEGMEG-----P 746
QY 950 FCKONNGCAKAKYKCSOKTQVS-----CCKKGYKGGDYSC-----TEIDPCA----- 993
Db 747 ICAQNTNDCSP--HPCVNSGTCTVDGDMWYRCECAPRAGP--DCRININECQSPCAFAT 803
QY 994 --DGVNG-GC-----HEHATCR-MTGRGKHKCECKSHYUGDV-----DEPEQLPRDR-- 1038
Db 804 CVDENIGYRCVCPRGHSGAKCEVSG--RPTTMGSVIPDGAKKDDNCTOCLNGRLA 860
QY 1039 -----CLQDNG--OCHPDASCA-----DLYFQDTTVGVFNHLSRPLGQYKLFPRDKAK 1082
Db 861 CCKVMCGRRPCLLHKHSGSPGSGSCIPLLDQCFVNRCTGVGECHSS-----SLQPVK 914
QY 1083 EACAKEA-----ATTA-TYN-QLSTQAKAYHLC 1110
Db 915 TKCTSDSYQDNCANITTFENKEMMSPLGTLTEHICS 950

RESULT 15
US-09-068-740A-7
; Sequence 7, Application US/09068740A
; Patent No. 6337387
; GENERAL INFORMATION:
; APPLICANT: SAKANO, SEIJI
; APPLICANT: ITOH, AKIRA
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE
; FILE REFERENCE: KP-8447
; CURRENT APPLICATION NUMBER: US/09/068,740A
; CURRENT FILING DATE: 1998-06-18
; PRIOR APPLICATION NUMBER: JP 7-299611
; PRIOR FILING DATE: 1995-11-17
; PRIOR APPLICATION NUMBER: JP 7-311811
; PRIOR FILING DATE: 1995-11-30
; PRIOR APPLICATION NUMBER: PCT/JP96/03356
; PRIOR FILING DATE: 1996-11-15
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1187
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-068-740A-7

Query Match 5.1%; Score 399.5; DB 4; Length 1187;
Best Local Similarity 21.7%; Pred. No. 3.8e-24;
Matches 229; Conservative 74; Mismatches 318; Indels 435; Gaps 68;

QY 189 POCVTRITTRACWLASLHNAKPARGEVYKMCALGTASVWDVNGTGTCCGSGFNGTACE 248
Db 196 PECNRAICRQGC-----SPKHGSKL-----PDCRCQYGMQGLYCD 232
QY 249 TCTEGKYGIHCQACSCVHGRCSQGPLGDSQCDVWGRVACDMEI-----TTDNC--NG 302
Db 233 KCLIPR-----GCVNGICTNE-----PMQCLCETNMGQGLCDMDLNCGRHQCCLNG 279
QY 303 TCHTSA-----NCLLDP-DGKASCK-----CAAGFRNG 330
Db 280 TCSNTPRDKYCCSEPEGYSGPCEIAEHNACLSDPCHNRRSGKETSIAGECECSPMTG-P 338
QY 331 TWCATAMCEISNGGCTSTADCKRTTPGRVRCVACAGYTG-----DGIVCLEINPLENH 385
Db 339 TCTSTNIDDCSPNN--CSHGCTQDLYNGRK-CVCPROMTGKTCQLDANEC-BAKPCV--- 391
QY 386 GGCDRNAECTGTGPNQAVCNCLPKYTGDGKVCSL-INVCITNNGGCSPPAFCNYTEODOR 444
Db 392 -----NAKCKNULIASYCDCLPQMMGN--CDININDCL--GCGQNDASGRDLVNGYR 441
QY 445 ICTCKRPDYTGDIYCRGSIYGBLPKNPSTSYFQLOEHAUYELAGRPFTYFAPLSSEF 504
Db 442 -CICPPGYAGDI--CERDI-DECAENPLN-----GHCQONEI---NRFQCLCPTGFSG 488
QY 505 N-----HEP-----RIKDW-----DOGGLMSQVLYRHVYGQQLLDNKK 539
Db 489 NLCQLDIDYCEBNPCONGAOCYNRASDYFCCKPREDEGNCNCHLKH---CR----- 537
QY 540 VTTSATTLTGGEBVSISVODTFVFINNEAKVLSSDIITNGVTHVIDKLSPNLLITPKD 599
Db 538 -TTPEVIDISCTVAAM-SNDT-----PEGVATISSNVCGR----- 570
QY 600 ALGRVLQNLTTVAANHGYTKFKSLIODSGLSVITDSIHTPTVTPMTDKALEALPREQO 659
Db 571 -----HG----- 572
QY 660 DFLFNQDNKDKLSYKLFHVITRDSKALASDLPRASAMKTLQSGSELSVRCTGSDIGELFL 719
Db 573 -----KCKSQSGKFTCCDNKG-----FT 591
QY 720 NEOMCRFIHGRGLFDYGAAYGIDCLLMPNLTGRCDTFTTFDIPBGCSGCIPTPKPLKS 779
Db 592 G-----TYCHENI-----NDC-ESNPCRNG-----GTCT----- 614
QY 780 KPRGVKK-KCIYNPLFRFRNVEGCONLTCTVVIOTPRCHGYMPDCA-----CPGGPDP 834
Db 615 --DGVNSYKCI-----CSDGMEGAYCETNINDC-----SQNP 644
QY 835 CNNRGMCRDLYTPMGQCLCHTGFNGTACELCMHGRGPRDQPRSCSEHGQC-DEG1TGS 893
Db 645 CHNGGTCRDLVNDP-YCDCKNGMKKCTC---HSR-DSQCDEATCNGGTCYDEG--DAF 696
QY 894 ECLCETGWTAAACDPTTAVFAVCTP-ACSVNATCTENN---TCVCNLNTEGSGITCTVVD 949
Db 697 KCMCGMGWGTTCN--IANNSSCLPRCHNGGTGVNNGSFTCYCKEGMEG-----P 746
QY 950 FCKONNGCAKAKYKCSOKTQVS-----CCKKGYKGGDYSC-----TEIDPCA----- 993
Db 747 ICAQNTNDCSP--HPCVNSGTCTVDGDMWYRCECAPRAGP--DCRININECQSPCAFAT 803
QY 994 --DGVNG-GC-----HEHATCR-MTGRGKHKCECKSHYUGDV-----DEPEQLPRDR-- 1038
Db 804 CVDENIGYRCVCPRGHSGAKCEVSG--RPTTMGSVIPDGAKKDDNCTOCLNGRLA 860
QY 1039 -----CLQDNG--OCHPDASCA-----DLYFQDTTVGVFNHLSRPLGQYKLFPRDKAK 1082
Db 861 CCKVMCGRRPCLLHKHSGSPGSGSCIPLLDQCFVNRCTGVGECHSS-----SLQPVK 914

Oy 1083 EACAKEA-----ATA-TYN-OLSYAKAKYHICS 1110
| : : | : : | : : | : :
Db 915 TKCTSDSYQDNCANITFTFNKEMMSPGLTTEHICS 950

Search completed: June 17, 2002, 12:29:43
Job time: 229 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:28:29 ; Search time 133.39 seconds
(without alignments)
3776.027 Million cell updates/sec

Title: US-09-842-930A-2
Perfect score: 7861
Sequence: 1 SLPSLLTRLRLQMPDYSIFRG.....WGHCGRPMRSGQATTVTPR 1431

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 3502263 seqs, 351980561 residues
Total number of hits satisfying chosen parameters: 3502263

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/2/paa/PCRTUS_COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US083_COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US084_COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US085_COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US086_COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US087_COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US088_COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US089_COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US090_COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US091_COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US092_COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US093_COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US094_COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US095_COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US096_COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US097_COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US098_COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US099_COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US100_COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US101_COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US102_COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US103_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	7861	100.0	1431	1	PCT-US01-13403-2 Sequence 2, Appl1
2	7861	100.0	1431	22	US-09-842-930A-2 Sequence 2, Appl1
3	5969	75.9	1394	1	PCT-US01-13403-25 Sequence 25, Appl1
4	5175.5	65.8	1192	1	PCT-US01-13403-62 Sequence 62, Appl1
5	5110	65.0	1394	22	US-09-842-930A-25 Sequence 25, Appl1
6	4906.5	62.4	1895	26	US-60-230-445-1388 Sequence 1388, Ap
7	4447.5	56.6	1069	1	PCT-US01-13403-63 Sequence 63, Appl1

8	3762.5	47.9	897	1	PCT-US01-13403-64 Sequence 64, Appl1
9	2813	35.8	2157	18	PCT-US99-30462-2 Sequence 2, Appl1
10	2813	35.8	2157	18	US-09-466-778-2 Sequence 2, Appl1
11	2810	35.7	1460	1	PCT-US01-13403-60 Sequence 60, Appl1
12	2797	35.6	1458	1	PCT-US01-13403-61 Sequence 61, Appl1
13	2710.5	34.5	669	21	US-09-715-417A-16 Sequence 16, Appl1
14	2701.5	34.4	669	21	US-09-715-417A-12 Sequence 12, Appl1
15	2131	27.1	1082	1	PCT-US01-16450-2725 Sequence 2725, Ap
16	2131	27.1	1082	1	PCT-US01-16450A-2725 Sequence 2725, Ap
17	1989.5	25.3	518	26	US-60-230-445-1786 Sequence 1786, Ap
18	1897.5	24.1	1039	26	US-60-207-214-329 Sequence 329, App
19	1881	23.9	1065	26	US-60-213-846-977 Sequence 977, App
20	1502.5	19.1	353	1	PCT-US99-30462-11 Sequence 11, Appl1
21	1502.5	19.1	353	18	US-09-466-778-11 Sequence 11, Appl1
22	1487	18.9	315	21	US-09-715-417A-4 Sequence 4, Appl1
23	1318.5	16.8	545	21	US-09-715-417A-18 Sequence 18, Appl1
24	1264.5	16.1	381	21	US-09-715-417A-14 Sequence 14, Appl1
25	1254.5	16.0	1535	26	US-60-230-445-1666 Sequence 1666, Ap
26	1209	15.4	244	21	US-09-715-417A-6 Sequence 6, Appl1
27	1186	15.1	1431	26	US-60-207-583-641 Sequence 641, App
28	1067	13.6	334	21	US-09-715-417A-20 Sequence 20, Appl1
29	1024	13.0	687	26	US-60-228-515-1080 Sequence 1080, Ap
30	774	9.8	457	1	PCT-US99-30462-5 Sequence 5, Appl1
31	774	9.8	457	18	US-09-466-778-5 Sequence 5, Appl1
32	762	9.7	457	16	US-09-244-112-108 Sequence 108, App
33	762	9.7	457	21	US-09-774-639-110 Sequence 110, App
34	762	9.7	457	23	US-09-969-730-110 Sequence 110, App
35	731.5	9.3	676	26	US-60-230-445-1625 Sequence 1625, Ap
36	685	8.7	835	26	US-60-230-445-1076 Sequence 1076, Ap
37	497.5	6.3	2871	19	US-09-538-092-1076 Sequence 1076, Ap
38	496.5	6.3	2911	22	US-09-825-751A-68 Sequence 68, Appl1
39	496.5	6.3	2912	1	PCT-US01-08631-36761 Sequence 36761, A
40	486	6.2	2523	12	PCT-US99-15817-3 Sequence 3, Appl1
41	486	6.2	2523	12	US-08-899-252-3 Sequence 3, Appl1
42	486	6.2	2523	15	US-09-121-457-3 Sequence 3, Appl1
43	484.5	6.2	2090	26	US-60-213-177-839 Sequence 839, App
44	484	6.2	2141	10	US-08-676-993-4 Sequence 4, Appl1
45	484	6.2	2141	10	US-08-676-993A-4 Sequence 4, Appl1

ALIGNMENTS

RESULT 1
PCT-US01-13403-2
Sequence 2, Application PC/TUS0113403
GENERAL INFORMATION:
APPLICANT: WEIGEL, PAUL H.
APPLICANT: ZHOU, BIN
APPLICANT: WEIGEL, JANET A.
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HYALURONAN RECEPTOR
FILE REFERENCE: 618754/JEP/199,538
CURRENT APPLICATION NUMBER: PCT/US01/13403
CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/245,320
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: 60/199,538
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 2
LENGTH: 1431
TYPE: PRT
ORGANISM: Rattus norvegicus
PCT-US01-13403-2

Query Match 100.0%; Score 7861; DB 1; Length 1431;
Best Local Similarity 100.0%; Pred No. 0;
Matches 1431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SLPSLLTRLRLQMPDYSIFRGYIIHYNLASAIESADAYTFVPNNEAIENYTRKKATSLK 60
|||||

```

Db 1 SEPSLLTRLEQMPDYSIFRGYIIHYNLASAIESADAYTVFVPPNNEAIENYIREKATSLK 60
OY 61 EDILRYHVVLGKELKNDLHNGMHRMTLMGFSYLLAFELRNDOLYVNEAPINTVATWK 120
Db 61 EDILRYHVVLGKELKNDLHNGMHRMTLMGFSYLLAFELRNDOLYVNEAPINTVATWK 120
OY 121 GVHGLEKYLEIÖKKNCDNDNTIIVREGCKGSQAAPCEPLETRPLRETRKCIYSIYFMK 180
Db 121 GVHGLEKYLEIÖKKNCDNDNTIIVREGCKGSQAAPCEPLETRPLRETRKCIYSIYFMK 180
OY 181 RSVFVIGCOPOCVRTITTRACWMLASLAHNAKPARGEYKMCALGTASVWDGVNGTGOCGL 240
Db 181 RSVFVIGCOPOCVRTITTRACWMLASLAHNAKPARGEYKMCALGTASVWDGVNGTGOCGL 240
OY 241 GENGTACECTEGBKYGIHCDQACSCYHGRCSOGPLDGSDCDDCVGRGKCDMEITTDNC 300
Db 241 GENGTACECTEGBKYGIHCDQACSCYHGRCSOGPLDGSDCDDCVGRGKCDMEITTDNC 300
OY 301 NGCTHTSANCLLDDPGKASCKCAGFRNGTGTCTAINACETSNGGCSTRADCKRTTPGNR 360
Db 301 NGCTHTSANCLLDDPGKASCKCAGFRNGTGTCTAINACETSNGGCSTRADCKRTTPGNR 360
OY 361 VCVCAAGTIGDGIIVCLEINPCLEHNGGCDRNAECTGTGNOAVCNCLPKYTGDKVCSLI 420
Db 361 VCVCAAGTIGDGIIVCLEINPCLEHNGGCDRNAECTGTGNOAVCNCLPKYTGDKVCSLI 420
OY 421 NVCCLNNGGCSPPAFANCNTYEDORICTCKPDYTGDIIVRGSIYGLPKNPSTSOYFOL 480
Db 421 NVCCLNNGGCSPPAFANCNTYEDORICTCKPDYTGDIIVRGSIYGLPKNPSTSOYFOL 480
OY 481 OEHAVERELAGPGPTVFAPLSSSFNHEPRIKDMDQOGLMSOVLRYHVNGCOQLLDNLKY 540
Db 481 OEHAVERELAGPGPTVFAPLSSSFNHEPRIKDMDQOGLMSOVLRYHVNGCOQLLDNLKY 540
OY 541 TTSATTTLOGEPPSISVSOTVFINNEAKYLSDDIISTNGVIHVIDLSPKNLLITPKDA 600
Db 541 TTSATTTLOGEPPSISVSOTVFINNEAKYLSDDIISTNGVIHVIDLSPKNLLITPKDA 600
OY 601 LGRVLONLTVAANHGTFTFSKLIODSGLLSVITDSIHPPVFWPMTDALEALPEEOD 660
Db 601 LGRVLONLTVAANHGTFTFSKLIODSGLLSVITDSIHPPVFWPMTDALEALPEEOD 660
OY 661 FLFNODNMDKLSYLFVHIRDSKALASDLPRASWKTLOGSELVRCGTGSDIGELFLIN 720
Db 661 FLFNODNMDKLSYLFVHIRDSKALASDLPRASWKTLOGSELVRCGTGSDIGELFLIN 720
OY 721 EÖMCRFIHGGLLFPDGVAVGIDCLMNPFLGRCDTFTTFDIPGEGSCSIFTPKCLKSK 780
Db 721 EÖMCRFIHGGLLFPDGVAVGIDCLMNPFLGRCDTFTTFDIPGEGSCSIFTPKCLKSK 780
OY 781 PKGVKKCIYNPLPFRANVEGCONLTCTVVIÖTPRCCHGFMPDQACPGSPPTPCNNRGM 840
Db 781 PKGVKKCIYNPLPFRANVEGCONLTCTVVIÖTPRCCHGFMPDQACPGSPPTPCNNRGM 840
OY 841 CSDLVTPMGÖCLCHNGFNGTACELCWHGREFPDQÖRSCSEHQÖCEGJITGSECLCEFG 900
Db 841 CSDLVTPMGÖCLCHNGFNGTACELCWHGREFPDQÖRSCSEHQÖCEGJITGSECLCEFG 900
OY 901 WTAASDPTFAVAVCTPACSVHATCTENNTCYCNLTNGBDGJITCTVVPFCCKNNNGGCAK 960
Db 901 WTAASDPTFAVAVCTPACSVHATCTENNTCYCNLTNGBDGJITCTVVPFCCKNNNGGCAK 960
OY 961 VAKCSÖKGÖVÖSCSKKGKGGDYSCTEIDPCADGVNGGCHHATÖRMÖPGKHKECKS 1020
Db 961 VAKCSÖKGÖVÖSCSKKGKGGDYSCTEIDPCADGVNGGCHHATÖRMÖPGKHKECKS 1020
OY 1021 HVVGSDVDEPEÖQLPDLRCLÖDNGÖCHPDASCADLYFÖDTTVGVFLRSPÖLQÖKYLTFPK 1080
Db 1021 HVVGSDVDEPEÖQLPDLRCLÖDNGÖCHPDASCADLYFÖDTTVGVFLRSPÖLQÖKYLTFPK 1080
OY 1081 AÖBACAKÖAATÖTÖTÖSTÖYÖKÖKÖYLÖCSÖGÖLESÖGRAVÖPTTÖYÖSÖKÖGANVÖIVÖY 1140
Db 1081 AÖBACAKÖAATÖTÖTÖSTÖYÖKÖKÖYLÖCSÖGÖLESÖGRAVÖPTTÖYÖSÖKÖGANVÖIVÖY 1140

```

```

OY 1141 GSRANKSEMDVFCYRMKDVNCTCKAGYVDGDFSGGNLÖVLMSPSTLNFLEVLAFS 1200
Db 1141 GSRANKSEMDVFCYRMKDVNCTCKAGYVDGDFSGGNLÖVLMSPSTLNFLEVLAFS 1200
OY 1201 KSSARGÖAFLKHLTÖDISÖTÖLFPÖNSGLPÖNKSÖLSDÖDIEHHLTNVAVSFYNDLVNCT 1260
Db 1201 KSSARGÖAFLKHLTÖDISÖTÖLFPÖNSGLPÖNKSÖLSDÖDIEHHLTNVAVSFYNDLVNCT 1260
OY 1261 FLRTMÖSÖLLTÖESÖDÖLHÖETRFYDGRSIIÖMDIÖANGIÖLHITÖSEPLRÖAPTÖATA 1320
Db 1261 FLRTMÖSÖLLTÖESÖDÖLHÖETRFYDGRSIIÖMDIÖANGIÖLHITÖSEPLRÖAPTÖATA 1320
OY 1321 HSGÖLÖTÖICÖAVÖLÖTÖGÖALÖALÖAÖSYÖFLÖKÖRTÖTÖGÖFÖRÖFÖDÖKÖRTÖLMSÖLLASSÖPRIÖSÖT 1380
Db 1321 HSGÖLÖTÖICÖAVÖLÖTÖGÖALÖALÖAÖSYÖFLÖKÖRTÖTÖGÖFÖRÖFÖDÖKÖRTÖLMSÖLLASSÖPRIÖSÖT 1380
OY 1381 LCMRÖPÖRRHÖPÖSPVPTÖSÖTÖENRÖTÖKÖRTÖLÖMÖHÖGÖPDMRSÖQÖATÖTVÖPR 1431
Db 1381 LCMRÖPÖRRHÖPÖSPVPTÖSÖTÖENRÖTÖKÖRTÖLÖMÖHÖGÖPDMRSÖQÖATÖTVÖPR 1431

RESULT 2
US-09-842-930A-2
; Sequence 2, Application US/09842930A
; GENERAL INFORMATION:
; APPLICANT: Weigel, Paul
; TITLE OF INVENTION: Identification of Hyaluronan Receptor for Endocytosis
; FILE REFERENCE: 5820.603
; CURRENT APPLICATION NUMBER: US/09/842,930A
; CURRENT FILING DATE: 2001-04-22
; PRIOR APPLICATION NUMBER: 60/245,320
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: 60/199,538
; PRIOR FILING DATE: 2000-04-25
; NUMBER OF SEQ ID NOS: 56
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 1431
; TYPE: prt
; ORGANISM: Rattus norvegicus
US-09-842-930A-2

Query Match 100.0%; Score 7861; DB 22; Length 1431;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1431; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```


Db 661 DAKVAVDLPTSTAMKTLQSGSELVSKGAGRIDGLFLNGQTCRIYQRELLFDLGAVYGI 720
OY 742 DCLAMNPTLGRCDPTFTTFFEDIPGEGSCITFTPKCKPLKSPGVKKKCYNPLPFRANEG 801
Db 721 DCLLIDPLGRCRDTFTTFFEDASGEGSCVNTSPCRPMKSPGVKOKCLYN-LPFRANEG 779
OY 802 CONLCTVVIQTPRCCGHGFMPCQACPGGPDTPCNRNRCMDLTYPMGQCLCHTFNGSTA 861
Db 780 CREGSLVITQIPROCKGVFRGDCQACPGGPAPCNRNGVCLDQYVSATGECCKNTGFNGTA 839
OY 862 CELCHHGFPGDCCPRSCSEHGQDEGTTGSEGLCEFTGWTAAASCDTPTAVAVCTPACS 921
Db 840 CEMCMPGRFGDCLPCCGSDHGQCDGTTGSGQCLCEFTGWTGTPSCDTPAVLPAVCTPCS 899
OY 922 VHAOTCTENNVCNLTNYGDDGTTCTVVPCKONNGCAKAVKCSKGTQVSCSKCKYKG 981
Db 900 AHATCKENNTBECNLDYTGDDGTTCTVVPCKODNGCAKAVARCSKGTQVSCSKCKYKG 959
OY 982 DGYSCIEIDPCADGVNGGCHHATCRMTGPKHKCECKSHYVGDGDEPEBOLPLDRCLQ 1041
Db 960 DGHSTETIDPCADGLNGGCHHATCKMTGPKHKCECKSHYVGDGLNCEPEQLPIDRCLQ 1019
OY 1042 DNGOCHPAPASCADLYFQDTTYGVFHLRSPLOYKLTLPDKAKACAKKEATITATYNOLSYA 1101
Db 1020 DNGOCHADAKCVDLHFQDTTYGVFHLRSPLOYKLTLPDKARACANEATMATYNOLSYA 1079
OY 1102 OKAKYHLCSAGMLSEGRVAYPTTYASOKGANVGVIVDYGSRANKSEMDVPCYMKDNV 1161
Db 1080 OKAKYHLCSAGMLSEGRVAYPTTYASOKGANVGVIVDYGSRANKSEMDVPCYMKDNV 1139
OY 1162 CTCRAGYVGDGFGSGNLLQVLMSPSLTNLTLEVLAFSKSSARQAOAFKLHLDLSIRCT 1221
Db 1140 CTCRAGYVGDGFGSGNLLQVLMSPSLTNLTLEVLAFSKSSARQAOAFKLHLDLSIRCT 1199
OY 1222 LEVPONSGLPKGSLSISGRDIEHLLTNVNVSEYNDLVNGFLTMTGSOGLITFEODOLH- 1280
Db 1200 LEVPONSGLPKGSLSISGRDIEHLLTNVNVSEYNDLVNGFLTMTGSOGLITFEODOLH- 1259
OY 1281 QETRVVDRSILQMDIILANGILHISEPLRAPPTAATAHSGLTGTFCAVAVLTGATA 1340
Db 1260 TETRVVDRSILQMDIILANGILHISEPLRAPPTAATAHSGLTGTFCAVAVLTGATA 1319
OY 1341 LAAYTFRLKORTTGFQRPDKRTMTSMLAS-----SSPRTISOTLCMRPQRHQPSP 1394
Db 1320 LAAYTFRLKORTTGFQRPDKRTMTSMLAS-----SSPRTISOTLCMRPQRHQPSP 1376
OY 1395 VPPSQ 1399
Db 1377 FTDSE 1381

RESULT 4
PCT-US01-13403-62
Sequence 62, Application PC/TUS0113403

GENERAL INFORMATION:
APPLICANT: WEIGEL, PAUL H.

APPLICANT: ZHOU, BIN

APPLICANT: WEIGEL, JANET A.

TITLE OF INVENTION: IDENTIFICATION AND USES OF A HYALURONAN RECEPTOR

FILE REFERENCE: 618754/JP/199,538

CURRENT APPLICATION NUMBER: PCT/US01/13403

CURRENT FILING DATE: 2001-04-25

PRIOR APPLICATION NUMBER: 60/245,320

PRIOR FILING DATE: 2000-11-02

PRIOR APPLICATION NUMBER: 60/199,538

PRIOR FILING DATE: 2000-04-25

NUMBER OF SEQ ID NOS: 64

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 62

LENGTH: 1192

TYPE: PRT

ORGANISM: Homo sapiens

PCT-US01-13403-62

Query Match 65.8%; Score 5175.5; DB 1; Length 1192;
Best Local Similarity 77.7%; Pred. No. 0;
Matches 914; Conservative 105; Mismatches 146; Indels 11; Gaps 4;
OY 231 NGTGTCCGGLGNGTACETCTGKGTGHCDDQACSCVHGRCSGPGLDSSCCDDGVGRVYK 290
Db 8 DGTGVCCEGEGFSGTACTCTEGKYGHCDDQACSCVHGRCSGPGLDSSCCDDGVGRVYK 67
OY 291 CDMELTDMNGTCHTSHANCLIDPDGKASCKCAAGFRNGTNGTCTAINTCETNSGCTSKA 350
Db 68 CDMATTEPNCGTCHTSHANCLIDPDGKASCKCAAGFRNGTNGTCTAINTCETNSGCTSKA 127
OY 351 DCKRTTPGNRVVCCKAGYTGDIYVCLINPCLENHGGCDRNAECTOTGPBNAVNCLEPRY 410
Db 128 DCKRTTPGNRVVCCKAGYTGDIYVCLINPCLENHGGCDRNAECTOTGPBNAVNCLEPRY 187
OY 411 TGDGVVCSLINVCLTNGGCSFPFACNTTEDQRICTCKPDTYGGIYCRGSIYELRKN 470
Db 188 TGDGVVCSLINVCLTNGGCSFPFACNTTEDQRICTCKPDTYGGIYCRGSIYELRKN 247
OY 471 PSTSQYFQLOEHAVERLAGPPTVFAPLSSFNHPEPRKMDQGLMSQVLRHYVVC 530
Db 248 PSTSQYFQLOEHAVERLAGPPTVFAPLSSFNHPEPRKMDQGLMSQVLRHYVVC 307
OY 531 QOLLDMNKVTTSATTLQGEPIVSVSDTVEINNEAKVLSDDIISTNGVHVHIDKLSP 590
Db 308 HOLLDMNKVTTSATTLQGEPIVSVSDTVEINNEAKVLSDDIISTNGVHVHIDKLSP 367
OY 591 KMLLTPKDAAGRYLONLTVAANHGYRKESKLDGSLIVTOSIHPVAVFVPTPKA 650
Db 368 KMLLTPKDAAGRYLONLTVAANHGYRKESKLDGSLIVTOSIHPVAVFVPTPKA 427
OY 651 LEALPEEODFLFNODNKKDKLSYLFHYIRDSKALASDLPSASWKTLOGSELSVRCGT 710
Db 428 LEALPEEODFLFNODNKKDKLSYLFHYIRDSKALASDLPSASWKTLOGSELSVRCGT 487
OY 711 GSDIGELFLNEOMCRPIHRLGLEDVGVAYIDCLLMNPTLGRCDPTFTTFFEDIPGEGSCIT 770
Db 488 GSDIGELFLNEOMCRPIHRLGLEDVGVAYIDCLLMNPTLGRCDPTFTTFFEDIPGEGSCIT 547
OY 771 FTTPCKPLSKPKGVKKKCYNPLPFRANVEGCONLTCTVVIQTPRCCGHGFMPCQACPG 830
Db 548 FTTPCKPLSKPKGVKKKCYNPLPFRANVEGCONLTCTVVIQTPRCCGHGFMPCQACPG 806
OY 831 PDPCCNRRGMDLTYPMGQCLCHTFNGTACELCHHGRFGDCCPRSCSEHGQDEGTT 890
Db 607 PDPCCNRRGMDLTYPMGQCLCHTFNGTACELCHHGRFGDCCPRSCSEHGQDEGTT 666
OY 891 GSSEGLCEFTGWTAAASCDTPTAVAVCTPACSVHATCTENNTCYCNLTNGDGTCTVVPF 950
Db 667 GSSEGLCEFTGWTAAASCDTPTAVAVCTPACSVHATCTENNTCYCNLTNGDGTCTVVPF 726
OY 951 CKONNGCAKAVKCSOKGTQVSCSKKYGKGDGYSCEIEDPCADGVNGGCHHATCRMTG 1010
Db 727 CKONNGCAKAVKCSOKGTQVSCSKKYGKGDGYSCEIEDPCADGVNGGCHHATCRMTG 786
OY 1011 PKHKECKSHYVGDGDEPEBOLPLDRCLQDNGQCHPDASCADLYFQDTTYGVFHLRSP 1070
Db 787 PKHKECKSHYVGDGDEPEBOLPLDRCLQDNGQCHPDASCADLYFQDTTYGVFHLRSP 846
OY 1071 LQGYKLTLPDKAKACAKKEATITATYNOLSYAKAKYHLCSAGMLSEGRVAYPTTYASQKC 1130
Db 847 LQGYKLTLPDKAKACAKKEATITATYNOLSYAKAKYHLCSAGMLSEGRVAYPTTYASQKC 906
OY 1131 GANVVGIVDYGSRANKSEMDVPCYMKDNVCTCKAGYVGDGFGSGNLLQVLMSPSLT 1190
Db 907 GANVVGIVDYGSRANKSEMDVPCYMKDNVCTCKAGYVGDGFGSGNLLQVLMSPSLT 966
OY 1191 NFLTEVLAFSKSSARQAOAFKLHLDLSIRGTLFVPPONSGILPGKKSLSGRDIEHLLTNVNV 1250
Db 967 NFLTEVLAFSKSSARQAOAFKLHLDLSIRGTLFVPPONSGILPGKKSLSGRDIEHLLTNVNV 1026

[illegible]

Query Match	Similarity	65.0%:	Score 5110:	DB 22:	Length 1394:
Best Local	Similarity	68.4%:	Pred. No. 0:		
Matches	951:	Conservative	124:	Mismatches 194:	Indels 122:
					Gaps 15:
QY	22	IIHYNLASIESADAYTVFVFNNEAIIENYIRPKKATSLKEDILRYHVYIGEEKLKNKDLHN	81		
Db	100	VIQYVNLAMNIEADAYTFAPRNNAIIENYIRKKVLSLEEDLRLRHVYIEKKLKNKDLHN	159		
QY	82	GMHRETMIGFSTLAFLELRNDQLYVNEAPIINTYNATDKGVI-HGLEKYLEIQKNR- C	137		
Db	160	GMHRETMIGFSTLAFLELRNDQLYVNEAPIINTYNATDKGVIQYVNLAMNIEADAYTFA	219		
QY	138	DNNDPI--IVRGECCKSQOAPCLETRKTRCISIIYPMGRKSVFTIGCOPOCVRTI	195		
Db	220	PNNNAIENYIR-----EKKVLSLEEDLR-----YHVYLEEK-----	251		
QY	196	ITRACIMLASIHNAPPARGEVYKMCALGTASVMDGVNNGTCCCGGFGFNACETGCEKY	255		
Db	252	-----LKNKDLHN-----GMR-----ETMIGFSTFL-----SF	275		
QY	256	GIHCDAQSCVHGRCSGQPLGDGSCDCDVGWRGVACDMEITTDNCGTCHTSANCLLDPD	315		
Db	276	FLHNQUL-----YVNEAPIINTYNATDKG-----	299		
QY	316	GKASCKCAAGFRGNQTVCTAINACETSNMGCSKADCKRTTPGNRVYCVKAGYTGDIYVC	375		
Db	300	-----VCAAGFGQNGTICTAINACEISNMGCSAKADCKRTTPGRRVCTCKAGYTGDIYVC	354		
QY	376	LEINPCLNHGGCDNNAECTGTGPMOAVCNCLPKYTGSGKVCSTLINVCLTNNGGSGSPFAF	435		
Db	335	LEINPCLNHGGCDNNAECTGTGPMOAVCNCLPATYTGDKVCTLLINVCLTNNGGSGSEFAI	414		
QY	436	CNYTEODORITCGKDYDYGDIYVCGSIYIGELLPKNPKSTSYQFFQLOEHNARELAGGPTT	495		
Db	415	CNHTEQVVERTCTCKRKNYIIGDGFTCGSIYQELPKNPKSTSYQFFQLOEHNAREKLVGPGPTT	474		
QY	496	VFAPLSSSFNHEPRIKMDQGLMSQVLYRHVVGCGQQLLDNLKVTTSATTLQGEVSVTS	555		

Db	475	VFAFLSAAPFDEEARVKOMDKYTGMLPQVLRHIVHACQJLLENNKLLSNATSLQGEPIYIS	534
OY	556	VSDOTVEINNEAKVLYSDIITNGCVIHVIDKLISPKNLLITERDALGRVLYLNTVAANH	615
Db	535	VSOSTVYIINNKAKLIISSDIISTNGCIYHIDKLISPKNLLITERDKMSGRIQMLTTLATNN	594
OY	616	GTYTFESKLIDDSGLIYITMSIHPTVTPVMPTRKALFALPREOOPFIENODKDKLSYL	675
Db	595	GYIFESMLIDDSGLIYITPIHTPTVLEPPTDQALHALPAEOOPLEFNODKDKLEYL	654
OY	676	KFHNIIDSKALASDUPKPSASAMKTLQOSELSVRCGTSGDIGELFLEDMOERFHRLDLDV	735
Db	655	KFHNIIDRAKVALDLPSTAMKTLQOSELSVYKGAARDIGDLFLNGQICRIYQRELLFL	714
OY	736	GVAATGIDCLLMNPPLGGRCDTFTTTFEDIPEGBCSCIETPKCPLSKRPKGVKKCIYNPPLF	795
Db	715	GVAATGIDCLLIDPLTIGRCDTFTTTPASGRCGCVNTPSCPRWMSRKGVKQCLYNP-LPF	773
OY	796	RRNVEGONLCTYVTOYPRCCCHGVFMPDQACGGGPDTPCNNRGMCRDLYTPMGQCLHT	855
Db	774	KRNLEGRERECISLYTOIPRCCCKQYFGCRDQACGGGADAPCNBRGVCLDQYSAITGECCKNT	833
OY	856	GFNGTAGELCMHGRFGPDCQPRSCSEHGQDEGICITGSGCLCETGWTATASCDTPPAVEAY	915
Db	834	GFNGTAGELCMWCPGRFGPDLPCGCSHDGQDDGITGSGQCLCETGWTGHCOTQVAILPAY	893
OY	916	CTPACSVHATCTENNTVCMLNTEBGGITCTVYVDFCKONNGGCAYAKCSQKGTQVSSC	975
Db	894	CTPACSAHATCKENNTCEMLIDYEGBGITCTVYVDFCKONGGCAYAVARCSQKGTVSSC	953
OY	976	KKGKYGAGYSCIEIDPCADGVNCGCHENHATRMGTGPKHKCECKSHYVDGVDCEPEOLP	1035
Db	954	QKGKYGAGHSTELIDPCADGLNCGCHENHATCKMTGPKHKCECKSHYVDGJLNCPEBOLP	1013
OY	1036	LDRCLODNGOCHPDASCADLYFODTTVGVFHLRSLPGQYKLPFDKAKEAKAEATYATY	1095
Db	1014	IDRCLODNGOCCHADAKCVDLHPDFTTVGVFHLRSLPGQYKLPFDKAREKANEATYMATY	1073
OY	1096	NOLSYAOKAKYHLCSAGMLESBGRVAAPTTYASQKGCANVGVIVDGSFRANKSEMDVDYCY	1155
Db	1074	NOLSYAOKAKYHLCSAGMLETGRKATPTAFASQNCSGVAVGIVDGPBPNNSEMDVDYCY	1133
OY	1156	RMKDVNCCTKAGYVGDGFSCSGMLLOYLWMSFPSTLTFLELYALFASKSSARGOAFKLHLLTD	1215
Db	1134	RMKDVNCCTKAGYVGDGFSCSGMLLOYLWMSFPSTLTFLELYALWSSSARGAFLHLLTD	1193
OY	1216	LSTNGTLFVPONSGLPCKNKLSCGRDIEHLLTNVWNSFYVDLYNGFFLRTMLGSQLITFS	1275
Db	1194	LSTNGTLFVPONSGLGENETLISGRDIEHLLTNVWNSMFPYNDLYNGTTQLRTGLSKSLITPAS	1253
OY	1276	ODOLH-OETRRVDOKRSILQWMDIIAANGLIHLIISEPLRAPPTATTAHSGDITGTCFCAVYL	1334
Db	1254	ODPLPPIETRRVDOKRALQWMDIFASKCIILHVISRPLKAPAPAYVTLTHRTGLGAGIEFALLI	1313
OY	1335	VTGAIALAAASYFLPKQORTTGFQSPDOKRTJMSMLAS-----SSPRIQTLTMRPQRR	1388
Db	1314	VTGAVALAAASYFLINRRTIGFQHFESSEEDINVAALGKQOPENTSNLPYESTTSAPE--	1371
OY	1389	HPOSPVTPPSQ	1399
Db	1372	-PSYDPEPTDSE	1381

```

RESULT      6
US-60-230-445-1388
: Sequence 1388, Application US/60230445
: GENERAL INFORMATION:
: APPLICANT: Beasley, Ellen
: TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
: TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN, TRANSPORTER PROTEINS,
: TITLE OF INVENTION: AND USES THEREOF
: FILE REFERENCE: CI000765

```

CURRENT APPLICATION NUMBER: US/60/230.445
CURRENT FILING DATE: 2000-09-06
NUMBER OF SEQ. ID NOS: 3051
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1388
LENGTH: 1895
TYPE: PRT
ORGANISM: HUMAN
US-60-230-445-1388

Query Match 62.4%; Score 4906.5; DB 26; Length 1895;
Best Local Similarity 65.6%; Pred. No. 0;
Matches 912; Conservative 105; Mismatches 158; Indels 215; Gaps 13;

QY 22 ITH-----YNLSAIESADATVPVNNENATENTIREKATSLKEDILRYVHVEKLLK 76
DB 696 VITHINKOYNLANAIEADATVPAPNNNAIENTIREKVLSDLEVDLRYHVLBEKLLK 755
QY 77 NDLEHGHRETMLGFSYLLAFELRNDOLYVNEAPINTVNAVTDKGVJHGLEKVEIQKRN 136
DB 756 NDLEHGHRETMLGFSYLLAFELRNDOLYVNEAPINTVNAVTDKGVJHGLEKVEIQKRN 815
QY 137 CDNDTIIIVGECKCSQAAPCPLETPKRLRETRCITISYHMKRSVFTGCOPOCVRTII 196
DB 816 CDNDTIIIVGECKCSQAAPCPLETPKRLRETRCITISYHMKRSVFTGCOPOCVRTII 857
QY 197 TRACMLSLANNAKPAPEVKMCAIGTASVMDVNGTGTCCGIGENFACETCEGKYG 256
DB 858 TRECCAGFFGQCPCEPNAONCVGNGICLDVYNGTGTCCGIGENFACETCEGKYG 917
QY 257 IHCDOASCYHGRCSQGPLDGSQCDVGMGRVXCDMEITTDNCGTCHTSANCLLPDG 316
DB 918 IHCDO----- 922
QY 317 KASCKCAAGFNGTVCATINACETSNNGCGSTKADCKRTTPGNRYCVKAGYTGDIYCL 376
DB 923 -----AINACEISNGCSAKADCKRTTPGNRYCVKAGYTGDIYCL 964
QY 377 EINPCLNHHGCDRNAECTGTGPNQAVNCPLRYTGDKVCSLIWCLTNNGGCSPPAFC 436
DB 965 EINPCLNHHGCDRNAECTGTGPNQ-----KNGCGSEPAFC 1000
QY 437 NYTEBODRICTCKPDYTGDIYVCRGSIYGLPKPNSTISQYFQLOEHAHRELACGPRTV 496
DB 1001 NYTEBODRICTCKPDYTGDIYVCRGSIYGLPKPNSTISQYFQLOEHAHRELACGPRTV 1060
QY 497 FAPLSSSFNHEPRIKDMQOGLMSQVLYHYVGOQLLDMLKVTSTATTOGEPVSTV 556
DB 1061 FAPLSSSFNHEPRIKDMQOGLMSQVLYHYVGOQLLDMLKVTSTATTOGEPVSTV 1120
QY 557 SQDVFVFINNEAKVLSDDITISNGVHVHDKLSPKNLITPKDALGRVLOMLTVAANH 616
DB 1121 SQ-----QNLTTLATNNG 1133
QY 617 YTKFSKLIODSGSLSVITDSIHPTVYFWPTDKALEALPREQODFLFNQDNKDKLSYK 676
DB 1134 YTKFSKLIODSGSLSVITDSIHPTVYFWPTDKALEALPREQODFLFNQDNKDKLSYK 1193
QY 677 FHVIRDSKALSDLPKRSKMTLQSGSELSVRCGTSIDIGELFLNQMGRFTHRGILPDVG 736
DB 1194 FHVIRDAKVALVADLPSTAMKTLOGSELVRCGAGRDIGDLFLNQTORIYORELLPDLG 1253
QY 737 VAYGIDCLLMPRTGLGRDFTTFPDIPGEGSCIFTPKCPKLSKSKRKKKKIYVLPFR 796
DB 1254 VAYGIDCLLMPRTGLGRDFTTFPD-----ASGVKOKCLYN-LPFR 1293
QY 797 RNVECCQNLCTVAVIOTPCCHGYEPMDCQACPGPDPCCNNKMGCRDLYTPMGQCLHTG 856
DB 1294 RNVECCQNLCTVAVIOTPCCHGYEPMDCQACPGPDPCCNNKMGCRDLYTPMGQCLHTG 1353
QY 857 FNGTACELCHMGRGFPDCCPNSCSHGQCCDSGITSSEGLCETGTAAASCDTPTAVFVNC 916
DB 1354 FNGTACELCHMGRGFPDCCPNSCSHGQCCDSGITSSEGLCETGTAAASCDTPTAVFVNC 916

DB 1354 FNGTA-----SC---GCSDHGQCDGDTGSGOCLCETGWTGSPCDTQ----AVC 1395
QY 917 TPACSVHATCTENNTVCNLYNEBGDITCTVVDCEKONNGCAKAVKCSQKGTGVSCSK 976
DB 1396 TPACSVHATCTENNTVCNLYNEBGDITCTVVDCEKONNGCAKAVKCSQKGTGVSCSK 1455
QY 977 KGYRGDGYSCIEIDPCADGVNGGCEHAHCTRMTPGKHCKECSKSHYVGDVCEPBDLP 1036
DB 1456 KGYRGDGYSCIEIDPCADGVNGGCEHAHCTRMTPGKHCKECSKSHYVGDVCEPBDLP 1515
QY 1037 DRCLQDNOGCHPADSCADLYRQDPTVGVFHLRSPLOGLYKLFEDAKKACAEATITATYN 1096
DB 1516 DRCLQDNOGCHPADSCADLYRQDPTVGVFHLRSPLOGLYKLFEDAKKACAEATITATYN 1575
QY 1097 QLSYAOKAKYHLCASGWLSEGRVAPPTVYASQKGANVAGIVDYGSRANKSEMDVFCYR 1156
DB 1576 QLSYAOKAKYHLCASGWLSEGRVAPPTVYASQKGANVAGIVDYGSRANKSEMDVFCYR 1635
QY 1157 MKDVNCTCKAGYVDGDESGCNLLQVLMSPSLNFLTLEVIAFKSSARGQAFKLHLDL 1216
DB 1636 MKDVNCTCKAGYVDGDESGCNLLQVLMSPSLNFLTLEVIAFKSSARGQAFKLHLDL 1695
QY 1217 SIRGTLFVPQNSGLPGNKSLSGRDIEHHLTVNVSFYVDLWNGTFLRTMLGSOLLTFPSQ 1276
DB 1696 SIRGTLFVPQNSGLPGNKSLSGRDIEHHLTVNVSFYVDLWNGTFLRTMLGSOLLTFPSQ 1755
QY 1277 DQHL-OETRFVYDGRSIILOMDIJAANGILHIISEPLRAPPTAATAHSGIAGTICAVLV 1335
DB 1756 DQHL-OETRFVYDGRSIILOMDIJAANGILHIISEPLRAPPTAATAHSGIAGTICAVLV 1815
QY 1336 TGAIALAAYSTFRLLKORTTGTQREDOKRTLMWLLAS-----SSPRISQTLCHRPQRRH 1389
DB 1816 TGAIALAAYSTFRLLKORTTGTQREDOKRTLMWLLAS-----SSPRISQTLCHRPQRRH 1872
QY 1390 POSPEVTPSQ 1399
DB 1873 PSYDPTDSE 1882
RESULT 7
PCT-US01-13403-63
Sequence 63, Application PC/TUS0113403
GENERAL INFORMATION: PAUL H.
APPLICANT: WEIGEL, PAUL H.
APPLICANT: ZHOU, BIN
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HVALUBONAN RECEPTOR
FILE REFERENCE: 618754/JP/199, 538
CURRENT APPLICATION NUMBER: PCT/US01/13403
PRIOR FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/245, 320
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: 60/199, 538
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 63
LENGTH: 1069
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-13403-63
Query Match 56.6%; Score 4447.5; DB 1; Length 1069;
Best Local Similarity 73.1%; Pred. No. 0;
Matches 798; Conservative 100; Mismatches 146; Indels 47; Gaps 6;
QY 316 GRASCKCAAGFNGTVCATINACETSNNGCGSTKADCKRTTPGNRYCVKAGYTGDIYV 375
DB 6 GRASCKCAAGFNGTVCATINACETSNNGCGSTKADCKRTTPGNRYCVKAGYTGDIYV 65
QY 376 LEINPCLNHHGCDRNAECTGTGPNQAVNCPLRYTGDKVCSLIWCLTNNGGCSPPAFC 435
DB 435 LEINPCLNHHGCDRNAECTGTGPNQAVNCPLRYTGDKVCSLIWCLTNNGGCSPPAFC 435

Db 66 LEINPCLENHGCGDKNAECTGPNQAAACNCLPAYTGDKVCTLLINVCILTKNGCSEPAI 125
QY 436 CNYTEODORICTCKPDYTGDIIVCGSIYIGELPKNPSTSOYFFOLQEHAVELAGPFT 495
Db 126 CNHGTGOVRICTCKPNYIGDGFCTGSGIYQELPKNPSTSOYFFOLQEHAVELAGPFT 185
QY 496 VFAPLSSSFNHPRIKMDQGLMSQVLRHYHVGCOQLLDNLKVTTSATTLQGPVSI 555
Db 186 VFAPLSSAFDEARVKMDKYGMLPQVLRHYHVAHQELLEMLKLSNATSLQGPVSI 245
QY 556 VSOOTVFNNNAKYLSSDITSTNGVIHIDKLSPKNLLITPKDLAGVLONLTTVAANH 615
Db 246 VSOOTVYNNNAKYLSSDITSTNGVIHIDKLSPKNLLITPKDLAGVLONLTTVAANH 305
QY 616 GYTFESKLIQSGSLVTTSDIHPFVTFWPTDKALEALPPQOQELFNQDNKDKLSYL 675
Db 306 GYTFESKLIQSGSLVTTSDIHPFVTFWPTDKALEALPPQOQELFNQDNKDKLSYL 365
QY 676 KFHVIKRSKALASDLPRSASMKTLQSGSLSVRCGTGSDIGELFLNQMCRPLHRLFLDV 735
Db 366 KFHVIKRSKALASDLPRSASMKTLQSGSLSVRCGTGSDIGELFLNQMCRPLHRLFLDV 425
QY 736 GVAAGIDCLLMNPTLGGCDFTFTFDIGECGSCITFTKPCPLKSPKVKKCIYNPLPF 795
Db 426 GVAAGIDCLLMNPTLGGCDFTFTFDIGECGSCITFTKPCPLKSPKVKKCIYNPLPF 484
QY 796 RRNEGCONLCTVVIOTPRCHGYFMPPCOACPGGPDTPCNRNRCGRDLTPMGOCLHT 855
Db 485 KRNEBGCERSLVIQIPRCKGYFGRDQOACPGGPDTPCNRNRCGRDLTPMGOCLHT 544
QY 856 GFNGTACELCMHGRFGPCOOPRSCSEHOGCEGTLTGSGECLEGTMTAASCDTPPAVAV 915
Db 545 GFNGTACELCMHGRFGPCOOPRSCSEHOGCEGTLTGSGECLEGTMTAASCDTPPAVAV 604
QY 916 CTPRACSVATCTENNTCYCNINTEBGTITCTVVPFCCKNONGCAKVAACSOQGTQVSSC 975
Db 605 CTPRACSVATCTENNTCYCNINTEBGTITCTVVPFCCKNONGCAKVAACSOQGTQVSSC 664
QY 976 KKGKGGDYSCEIDPCADGVNGGCEHATCRMTGPGKHKCEKSHYVGDDVDCPEOLP 1035
Db 665 OKGTGKGDSHSTETIDPCADGVNGGCEHATCRMTGPGKHKCEKSHYVGDDVDCPEOLP 724
QY 1036 LDRCLQDNGQCHPRASCADLYFODTTVGVFHLRSPLOQKLTLPDKAKACAEATTIATY 1095
Db 725 LDRCLQDNGQCHPRASCADLYFODTTVGVFHLRSPLOQKLTLPDKAKACAEATTIATY 784
QY 1096 NOLSYAOKAKYHILCSAGMLESGRAVAPPTTYASOKCANVGLVGYGSRANKSEMDVFCY 1155
Db 785 NOLSYAOKAKYHILCSAGMLESGRAVAPPTTYASOKCANVGLVGYGSRANKSEMDVFCY 844
QY 1156 RMKVUNCTCKAGYVGDGSCSGNLLQVLMSPSLTNFLTEVLAFSSKARQOAFJXHLTD 1215
Db 845 RMKVUNCTCKAGYVGDGSCSGNLLQVLMSPSLTNFLTEVLAFSSKARQOAFJXHLTD 868
QY 1216 LSIKGTLEFVNPNOSGLPGKNSLSGRDIEHLLTNVNVSEFYNDLVNGTFLRTMLGSQLITFS 1275
Db 869 LSIKGTLEFVNPNOSGLPGKNSLSGRDIEHLLTNVNVSEFYNDLVNGTFLRTMLGSQLITFS 928
QY 1276 OQOLH-QETRVVDGRSLIOMOLIAANGILHTISEPLRARPPLAATAHSGLTGTICAVYL 1334
Db 929 OQOLH-QETRVVDGRSLIOMOLIAANGILHTISEPLRARPPLAATAHSGLTGTICAVYL 988
QY 1335 VTGAIALAAAYFRLKORTTFORFODKRTLMSTMULAS-----SSPRISOTLCMRPQR 1388
Db 989 VTGAIALAAAYFRLKORTTFORFODKRTLMSTMULAS-----SSPRISOTLCMRPQR 1046
QY 1389 HPOSPVTPSQ 1399
Db 1047 -PSYDPFDTSE 1056

RESULT 8
PCT-US01-13403-64

Sequence 64, Application PC/TU0S0113403
GENERAL INFORMATION:
APPLICANT: WEIGEL, PAUL H.
APPLICANT: ZHOU, BIN
APPLICANT: WEIGEL, JANET A.
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HVALUROMAN RECEPTOR
FILE REFERENCE: 618754/Jp/199,538
CURRENT APPLICATION NUMBER: PCT/US01/13403
CURRENT FILING DATE: 2001-04-25
PRIOR APPLICATION NUMBER: 60/245,320
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: 60/199,538
PRIOR FILING DATE: 2000-04-25
NUMBER OF SEQ ID NOS: 64
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 64
LENGTH: 897
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-13403-64

Query Match 47.9%, Score 3762.5; DB 1; Length 897;
Best Local Similarity 76.1%; Pred. No. 6,6e-270;
Matches 676; Conservative 87; Mismatches 114; Indels 11; Gaps 4;

QY 519 MSQVLRHYHVGCGQQLLDNLKVTTSATTLQGPVSISSQDTVFNNNAKYLSSDITSTN 578
Db 1 MPOVLRHYHVAHQELLEMLKLSNATSLQGPVSISSQDTVFNNNAKYLSSDITSTN 60
QY 579 GVHVIKLSKPNLLITPKDLAGVLONLTTVAANHGYTFKPSKLIQSGSLVTTSDI 638
Db 61 GVHVIKLSKPNLLITPKDLAGVLONLTTVAANHGYTFKPSKLIQSGSLVTTSDI 120
QY 639 TPVTFWPTDKALEALPPQOQELFNQDNKDKLSYLFKHVIRDSKALASDLPRSASMK 698
Db 121 TPVTFWPTDKALEALPPQOQELFNQDNKDKLSYLFKHVIRDSKALASDLPRSASMK 180
QY 699 LQSGSLSVRCGTGSDIGELFLNQMCRPLHRLFLDVGAAGIDCLLMNPTLGGCDFT 758
Db 181 LQSGSLSVRCGTGSDIGELFLNQMCRPLHRLFLDVGAAGIDCLLMNPTLGGCDFT 240
QY 759 TFDIGECGSCITFTKPCPLKSPKVKKCIYNPLPFRRNEGCONLCTVVIOTPRCHG 818
Db 241 TFDIGECGSCITFTKPCPLKSPKVKKCIYNPLPFRRNEGCONLCTVVIOTPRCHG 299
QY 819 YFMPDQOACPGGPDTPCNRNRCGRDLTPMGOCLHTGFGNTACELCMHGRFGPCOOPR 878
Db 300 YFMPDQOACPGGPDTPCNRNRCGRDLTPMGOCLHTGFGNTACELCMHGRFGPCOOPR 359
QY 879 CSEHOGCEGTLTGSGECLEGTMTAASCDTPPAVAVCTPRACSVATCTENNTCYCNIN 938
Db 360 CSEHOGCEGTLTGSGECLEGTMTAASCDTPPAVAVCTPRACSVATCTENNTCYCNIN 419
QY 939 EGDGITTCTVVPFCCKNONGCAKVAACSOQGTQVSSCCKGKGGDYSCEIDPCADGVNG 998
Db 420 EGDGITTCTVVPFCCKNONGCAKVAACSOQGTQVSSCCKGKGGDYSCEIDPCADGVNG 479
QY 999 GCEHATCRMTGPGKHKCEKSHYVGDDVDCPEOLPLDRCLQDNGQCHPRASCADLYFQ 1058
Db 480 GCEHATCRMTGPGKHKCEKSHYVGDDVDCPEOLPLDRCLQDNGQCHPRASCADLYFQ 539
QY 1059 DTTVGVFHLRSPLOQKLTLPDKAKACAEATTIATYNOLSYAOKAKYHILCSAGMLESGR 1118
Db 540 DTTVGVFHLRSPLOQKLTLPDKAKACAEATTIATYNOLSYAOKAKYHILCSAGMLESGR 599
QY 1119 VAYPTTYASOKCANVGLVGYGSRANKSEMDVFCYEMKDVNCTCKAGYVGDGSCSGN 1178
Db 600 VAYPTTYASOKCANVGLVGYGSRANKSEMDVFCYEMKDVNCTCKAGYVGDGSCSGN 659
QY 1179 LLQVLMSPSLTNFLTEVLAFSSKARQOAFJXHLTDLSIKGTLEFVNPNOSGLPGKNSLSG 1238
Db 660 LLQVLMSPSLTNFLTEVLAFSSKARQOAFJXHLTDLSIKGTLEFVNPNOSGLPGKNSLSG 719

NUMBER OF SEQ ID NOS: 37
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 2
LENGTH: 2157
TYPE: PRT
ORGANISM: Homo sapiens
US-09-466-778-2

Query Match 35.8%; Score 2813; DB 18; Length 2157;
Best Local Similarity 40.1%; Pred. No. 7.8e-199;
Matches 562; Conservative 220; Mismatches 554; Indels 66; Gaps 20;

```
QY 5 LTRLEQMPDYSIFRGYIIHYNLASAIESADAYTFVFNNEIENYIREKATSLIKEDIL 64
DB LLOQDLVPAISLRELLQHNOLVQIEATYATITFVFNRSLE---AOGNSHLDADTV 774
QY 65 RYHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLYVNEAPINTVATDKGYIH 124
DB RHVVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLYVNEAPINTVATDKGYIH 834
QY 125 GLEKYLEIQKRNCDNDITTYRGECKSCQAAPC---PLETKPLRETRKCIYIYFMGK 180
DB GLSGVLIVGSSRCLSHAEALREKCVNCTRRFCTQGFOLDPTP---RKSCYVRSGFSFS 891
QY 181 RSVFICGPOCVRTITRACWMLASLAHNAKPARCEVKMCALGTASVMDGVNGTGCOCGL 240
DB 892 R-----GCTYTAKKIQVDDCCPFGFGTICPCPGGLGVCVSGHCCQDRLGSGRCHHE 947
QY 241 GPNGTACETEGKYGIIHCDQASCVHGRCSQGLGSDCDDCVMGRCVCKDMEITTDNC 300
DB 948 GFHGACAEVCEIGRYGPNCTGVCDAHGLCEGLGSDGSCVNCVMQGLRCDQKITSPOC 1007
QY 301 NGTCHTSANCLLDPRGKASCCKAAGFRNGYVCTAIIACETSNGGCTKADCKRTTPGKR 360
DB 1008 PRKCDPMANCVDSAGASTACAGYSGNGIFCSFVDPACAHGCGCSBHAMCTKYAPQGR 1067
QY 361 VCVCAGATYGTGIVCLLETINPCLEHNGCDRNAECTOTGPNOAVCDEKPYTGDG-KVCSL 419
DB 1068 TCTGODGYMGDELCOELINSLIHNGGCHIHAECTIPTGPQOYSCSCREYSTSDGIRTEEL 1127
QY 420 INVCITNNGGSPFAFCVTEODRITCTCKPDTY-GDGIVCRSITYGELPKNPSTSOYFF 478
DB 1128 LDPCSKNNGCCSPYATCKSTGDGRTCTDPAHTVGDGLTCARAVGDELLRDKHAS--FF 1185
QY 479 QLOEHAVNELAGPFGTYFAP---LSSSFNHE--PRIDMDQOGLMSOYLRIHVYVCCOL 533
DB 1186 SLRLLEYKELGDDGPTTFVPHADLMSLSODELARIRAHRO----LVFRYHVVGCRRLL 1240
QY 534 LIDNLKVTTSATTOGEFVSISSODTFEINNEAKVLSDDIISTNGVLIHVLDIKLSPKNL 593
DB 1241 RSEDLLEGVATLALSGHRLRSEREGSITYLNDFAVYSSDHEAVVGIILHFDRLVLLPEEA 1300
QY 594 LITRKDALGRVLQNTTVAAHNGYTKFSKLIDSGSLSVITDSIHTPTVTPMPTDKALEA 653
DB 1301 LHMEPPDAPITPRRNVYTAAGGFGYKIFSGLLKVLGALLPLRLREASHRPMTMLPPTDAFRA 1360
QY 654 LPREQODFLFNQDKDKLXYLTKFHVIRDSKALASDLPRASMKTLQOSELSVRCGTSD 713
DB 1361 LPPROAMLYHEDHDKDAALIRGMIRNVALASDLPLNLPLRTMHGTPIISFSC-SRTR 1419
QY 714 IGEFLFLEOMCRFIHRIIGLLPVGVAAGIDCLIMNPTLGRDPTTFPIIP-GECCSCLETF 772
DB 1420 PGLMVGDDOARIVORHLPFEGGLAYGIDQLLEPPGIGARCDHFETRLRLNTGTCIGLE 1479
QY 773 PKCPKLSKPKGVKKKC-----IYNLPFR---RNV-----EGCONLCITV 808
DB 1480 PPECEGSGEGSPACMRFYKFTSPRLSHLSGLRSVWVHSLMGRPOGLGRCHRNCTV 1539
QY 809 VIOFPRCHGHEMPDQACPGGPDTPCNRNMGMDLITPMDQOCLCHTFNGTAGLCHMG 868
DB 1540 TYMKPSCCPGHYGECCQACPGGSPSCSDKGVCMGMSGSOCLCRSGFAGTACELAPG 1599
```

```
QY 869 RFPDQPRSCSEHGOCDEGITGSGECLCEGTWTAASCDTPPAVAVCTPACSVHATCTE 928
DB 1600 AFGPHQACRCYHGHGCDLGGSSGCPDEBGTGRCEVOLELDQVCTPPCAFEAVCSA 1659
QY 929 NNTVCNLMYEGDGIYTCVDEFCQKNNNGCAKAVAKCSQKQOVSCSKGKQDGYSCIE 988
DB 1660 GNSCECSLGEEDGRVCTVADLCQDGHGSCSEHANSQYGTWYTCILPDYEDGMSCSA 1719
QY 989 IDPCADGVNGGCHENATCTMTGPCKHCECKSHYVDGYDC-EPEQDPLDRLQDNGQCH 1047
DB 1720 RNPCTDGHNGCSEHANCSTGLNTRRCEHAGYVDGQCLEESPEPPVDRCLGQPPCH 1779
QY 1048 PASCADLYEOPDTGVFPLRSPDLGOYKLTIPKAKCAKKAATATYNDLSAOKAKXH 1107
DB 1780 SDAMKTDLHFOEKRAVCFHLOATSGFYGLNFSBAEACACAGCVLASFPOLSAHDOUGFH 1839
QY 1108 LCSAGMLESGRVAYPTTYASQKGANVVGIVYGSFRANKSEMDVFCYBMKDVNCTCKAG 1167
DB 1840 LCLMGWLANGSTAHNPVFPVADCGNCRGIVSLGARKNLSEKMDACFPVQVADACRNG 1899
QY 1168 YVGDFG-CSGMLQVMSFSLTNFLTEVLAFSKSSARGQAFKHLTDLSTRTLEFPQ 1226
DB 1900 FVGDGISTCNGKLLDVLATANFSTFYGMLLGYANATQGLDFLDDELTYKTFLEPV 1959
QY 1227 NSGLQNKSLSGRDIENHLTNVNSFYNDLVNGTFLRTMLGSOOLLITFSQDQLHOTRFRV 1286
DB 1960 NEGFDNMFTLSGPNLHLASNAITLSAN-ASQGLLPAHSGSLITISDGPNNSSAPYA 2018
QY 1287 DG-----RSLQWDIIAANGILHIIEPLRAPP-----TAATAHSGLGIGFCVAVLYTG 1337
DB 2019 PGTVVVSRILTVMDIAPNGIHALASPLAPPQPAVILAXEAPVPAAGVAVLAAGALLG 2078
QY 1338 AIALAAYSFRLKORTTGFQRF 1359
DB 2079 LVAGALYLRARGKPMGFGFSAP 2100
```

RESULT 11
PCT-US01-13403-60
Sequence 60, Application PC/TUS0113403
GENERAL INFORMATION:
APPLICANT: WEIGEL, PAUL H.
APPLICANT: ZHOU, BIN
APPLICANT: WEIGEL, JANET A.
TITLE OF INVENTION: IDENTIFICATION AND USES OF A HVALURONAN RECEPTOR
FILE REFERENCE: 618754/Jp/199,538
CURRENT FILING DATE: 2001-04-25
PCT/US01/13403
PRIOR FILING DATE: 2000-11-02
PRIOR APPLICATION NUMBER: 60/7245,320
PRIOR FILING DATE: 2000-04-25
PRIOR APPLICATION NUMBER: 60/7199,538
NUMBER OF SEQ ID NOS: 64
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 60
LENGTH: 1460
TYPE: PRT
ORGANISM: Homo sapiens
PCT-US01-13403-60

Query Match 35.7%; Score 2810; DB 1; Length 1460;
Best Local Similarity 40.0%; Pred. No. 7.5e-199;
Matches 561; Conservative 222; Mismatches 546; Indels 72; Gaps 22;

```
QY 5 LTRLEQMPDYSIFRGYIIHYNLASAIESADAYTFVFNNEIENYIREKATSLIKEDIL 64
DB LLOQDLVPAISLRELLQHNOLVQIEATYATITFVFNRSLE---AOGNSHLDADTV 77
QY 65 RYHVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLYVNEAPINTVATDKGYIH 124
DB RHVVVLGKLLKNDLHNGMHRETMLGFSYLLAFELRNDOLYVNEAPINTVATDKGYIH 137
```

OY	125	GLKVEI	QKON	RDNNDT	II	VEGEGK	SOQA	PC	---	PLETKP	LEBTK	QTS	IY	PMK	180																																											
OY	126	II	II	II	II	II	II	II	II	II	II	II	II	II	181																																											
Db	138	GLSGLV	LIVGSS	RCLSH	HALE	REKCYM	TRR	RC	QT	Q	Q	Q	Q	Q	182																																											
OY	181	RSVFIC	QOPCV	VTII	IT	TRACW	LAS	LAHN	KAP	AGE	KNK	AL	GTAS	VMD	VGNGT	240																																										
OY	195	R	---	GC	SY	CAK	KI	QV	PO	CC	FE	GT	LE	PC	PG	L	GGV	SGH	Q	CC	DR	L	GS	GE	CH	E	250																															
Db	241	GFNG	STAC	ET	BE	KY	GI	HC	DO	AC	SV	HC	SG	PR	L	GD	SC	DD	VG	MG	RG	VC	K	DM	L	IT	DN	C	300																													
OY	251	GFHSTAC	VE	EL	RG	YR	PN	CT	GV	DC	HA	HC	L	OD	GL	GD	DS	CC	VN	MG	Q	L	R	DO	K	IT	SP	C	310																													
Db	301	NGCH	TS	AN	CL	LD	DD	KA	SK	AA	GR	RG	MG	Y	CTA	I	N	C	ET	SN	GG	CC	SK	AK	CT	ET	PG	N	360																													
OY	311	PRKCD	PAN	CV	Q	DS	AG	ST	OC	AC	AG	IS	GN	GT	CE	ND	PA	HC	GG	CC	PH	AN	CT	VA	PG	R	370																															
Db	361	VCVAC	AY	TD	GD	J	Y	C	L	E	I	N	C	L	E	N	H	GC	DR	N	A	E	CT	Q	T	GN	AV	NC	L	E	PK	Y	GD	E	K	Y	CS	L	419																			
Db	371	TC	TC	Q	D	S	T	C	O	D	S	M	D	G	E	L	C	O	E	I	N	S	C	L	I	H	H	G	G	C	I	I	A	E	C	P	T	R	P	O	Y	S	C	S	C	R	G	S	Y	S	G	D	I	R	C	E	L	430
OY	420	INVC	L	T	N	N	G	G	C	S	P	A	N	C	T	Y	E	O	D	R	I	C	T	K	P	D	Y	T	-	GD	I	V	C	R	G	S	I	Y	G	L	P	K	N	P	S	O	Y	F	478									
Db	431	LD	P	C	K	N	G	G	C	S	P	A	N	C	S	T	G	D	G	O	P	R	C	T	D	T	A	H	T	VG	D	G	L	T	C	A	R	V	G	E	L	L	R	D	K	H	A	S	-	FF	488							
OY	479	Q	LO	HA	V	R	E	L	A	G	R	P	T	V	A	P	---	L	S	S	T	N	H	E	-	P	R	I	D	M	O	O	G	L	M	S	O	V	L	R	H	N	V	G	O	L	533											
Db	489	S	L	R	L	E	K	E	L	E	K	G	O	P	F	I	F	P	H	A	D	L	M	S	N	I	S	O	D	E	L	A	R	H	R	N	O	---	L	V	F	R	Y	H	V	A	G	R	L	543								
OY	534	L	D	N	L	K	T	T	S	A	T	T	L	O	G	E	P	V	S	I	S	O	D	T	F	E	N	N	E	A	K	L	S	D	I	S	T	N	G	V	I	N	H	I	D	K	L	S	P	N	L	593						
Db	544	R	S	E	D	L	E	O	G	A	T	A	L	S	H	I	P	L	F	E	R	E	S	I	T	L	N	D	P	A	R	V	S	S	D	H	A	V	N	G	I	I	H	F	I	D	R	V	L	P	E	A	603					
OY	594	L	I	T	E	K	D	A	L	G	R	V	L	O	N	L	T	T	V	A	N	H	G	Y	T	F	E	S	K	L	I	O	D	S	L	I	V	T	D	S	I	H	T	P	V	T	E	M	P	D	K	L	E	A	653			
Db	604	L	H	M	E	R	D	A	P	I	R	R	N	V	T	A	A	G	G	F	E	K	T	F	S	G	L	I	K	V	A	G	L	L	P	L	R	E	A	S	H	R	P	T	M	P	D	A	F	R	A	6						

[illegible]


```
; APPLICANT: VerneL, Corine
; APPLICANT: Fernandes, Elma
; TITLE OF INVENTION: NOVEL POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
; FILE REFERENCE: 15966-606
; CURRENT APPLICATION NUMBER: US/09/715,417A
; CURRENT FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: 60/166,336
; PRIOR FILING DATE: 1999-11-19
; PRIOR APPLICATION NUMBER: 60/167,785
; PRIOR FILING DATE: 1999-11-29
; PRIOR APPLICATION NUMBER: 60/187,844
; PRIOR FILING DATE: 2000-03-08
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 12
; LENGTH: 669
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-715-417A-12
```

```
Query Match          34.4%: Score 2701.5; DB 21: Length 669;
Best Local Similarity 78.6%: Pred. No. 2.9e-191;
Matches 470: Conservative 60; Mismatches 67; Indels 1; Gaps 1;
```

```
QY 509 RIKMDQGLMSQVLYRYHVYVGCQQLLDNLKVTTSATTLQGEVSYSDTFEINNEAK 568
      ::::: ||| ||||| ||| ||||| ||| ||||| ::::: ||| ||||| ||| |||||
Db 68 QVKMDKYGCLMPQLRYHVYVACHOLLENLKLSNATSLQGEPIVISTOSQTVIINNAK 127

QY 569 VLSDIISTNGVIVHIDKLSLSPKMLITPKDALGRVLQNLTTVAANHGYTFESKLIODSG 628
      ::::: ||| ||||| ||| ||||| ||| ||||| ::::: ||| ||||| ||| |||||
Db 128 IISDIISTNGIIVHIDKLSLSPKMLITPKDNGSRILQNLTTLATNNGYIKFSNLIODSG 187

QY 629 LLSVITSIHPPVAVFVPTDQALALPEODDLFNQDNKMLKSYLKEHVIIRDSKALAS 688
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 188 LLSVITPIHPPVAVFVPTDQALALPEODDLFNQDNKMLKSYLKEHVIIRDSKALAV 247

QY 689 DLPRASMKTLQSGSELSVRCGTSGDIDELFLNEQMCRFIHRGLLEDVGVAYGIDCLLMP 748
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 248 DLPTAMKTLQSGSELSVRCGAGRDIQDLFLNGQTCRIYQELLFLDGVAIGIDCLLIDP 307

QY 749 TLGRCTFTFTFDLPGECSGCIPTPKCPKLPKGVKKCLLYNPLPRRNVGCONCLTV 808
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 308 TLGRCTFTFTFDASGEGSCVNTSPCRMSKPGVKKCLLYN-LPFRNLGGERRCSL 366

QY 809 VIOPRCHGTFMPDCOACPGPPTPCNNRGMCRDLTYPMGQCLCHTGFNGTACELCWG 868
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 367 VIOIPRCCKGTFGRDCOACPGGPPVAPCNRNGVCLDQISATGECCKNTGFTACEMCPG 426

QY 869 RFGPDQCPRSCSEHGQCDGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTE 928
      ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 427 RFGPDCLPRCGSDHGQCDGITYSGGQCLCETGWTGSPCDTQAVLSAVCTPCCSAHATCKE 486

QY 929 NNTCVMLNTEBGDITITCVNDFCKQNNGGCAKVAKSGKGTQVSCSKGKGIGSYCTE 988
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 487 NNTCEMLNTEBGDITITCVNDFCKQNNGGCAKVAKSGKGTQVSCSKGKGIGSHSCTE 546

QY 989 IDPCADVNGGCEHATCRMTGPGKHKECKSHYVGVGDVDEPOLPLDRCLQDNGOCHP 1048
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 547 IDPCADGNGGCEHATCRMTGPGKHKECKSHYVGVGDLEPDLPLDRCLQDNGOCHA 606

QY 1049 DASCADLYFDQTTGVFHLRSLPGLQYKLLTPDKAEACAKAEATATATYNQLSYAOKAKY 1106
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 607 DAKVDLHFDQTTGVFHLRSLPGLQYKLLTPDKAREACANEATMATYNQLSYAKTWT 664
```

```
RESULT
PCT-US01-16450-2725
; Sequence 2725, Application PC/TUS0116450
; GENERAL INFORMATION:
; APPLICANT: Human Genome Sciences, Inc.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA131PCT
```

```
; CURRENT APPLICATION NUMBER: PCT/US01/16450
; CURRENT FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: 60/205,515
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 2820
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2725
; LENGTH: 1082
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US01-16450-2725
```

```
Query Match          27.1%: Score 2131; DB 1: Length 1082;
Best Local Similarity 41.6%: Pred. No. 1.4e-148;
Matches 430: Conservative 155; Mismatches 396; Indels 52; Gaps 16;
```

```
QY 370 GDGIYCLEINPCLNHHGGCDRNAECTGTGPMQAVNCNLPKTYGQ-KVCSLINVCITNNG 428
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 2 GDGELCOEINSLIHHGGCHIHAEISTGPQVSCSREGSGDGIPTCELLDPCSKNNG 61

QY 429 GCSPPAFCNVTEQDQRICTCKPDYT-GDGIYCRGSISGELPKNPSTSQYFQLOEHAVRE 487
      ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 62 GCSPPATCKSTGDCQRTCTDTHFTVGDGLTCRARVGLRLDKHAS--FSLRLLEYKE 119

QY 488 LAGGPPTVFAP---LSSSFNHE--PRIKMDQGLMSQVLYRYHVYVGCQQLLDNLKVTT 542
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 120 LKGGPPTIIFVPHADIMSNLSQDELARIRAHKQ-----LVFRYHVYVGCRLRSDDLLEG 174

QY 543 SATTLQGEVSYSDTFEINNEAKVLSDIISTNGVIVHIDKLSLSPKMLITPKDALG 602
      ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 175 YATLASHPLHFRSEREGIYLNDFARVVSDHEAVNGILHFIIDVLLPPEALHWEPPDAP 234

QY 603 RVLQNLTTVAANHGYTFESKLIODSGILSVTDSIHPPVAVFVPTDQALALPEODDL 662
      ::::: ||| ||||| ||| ||||| ||| ||||| ::::: ||| ||||| ||| |||||
Db 235 IPRKNVIAAAGFGKLTFSGLKVAGLPLRLREASHRFTLMPDLDAFRALPPDQAWL 294

QY 663 FNQDNKMLKSYLKEHVIIRDSKALASDLPRASMKTLQSGSELSVRCGTSGDIDELFLNEQ 722
      ::::: ||| ||||| ||| ||||| ||| ||||| ::::: ||| ||||| ||| |||||
Db 295 YHEDHROKLAIRLGHMIRNVEALASDLPLNGLPTMGTPTFSFC--SRTPGELAMGED 353

QY 723 MCRFIHRGLLEDVGVAYGIDCLLMPNPTLIGRCDTFTTDPD-GECSGCIPTPKCPKLPK 781
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 354 DARIVORHLPEGGVLAIDQLLEPPGIGARCDHFETRLKNTCSIGLEPPCPDESQE 413

QY 782 KGVKKKC-----LYNPLPFR-----RNV-----EGCONCLTVVIOIPRCH 817
      ::: ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 414 QGSPFACRFTPKFTWTSPLHSLGIRSVWHPSLMGRPOGIGRCHRCVTTTMMKPSCP 473

QY 818 GFMPDCOACPGPPTPCNNRGMCRDLTYPMGQCLCHTGFNGTACELCWGRFGPDCQPR 877
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 474 GHVSECCACPGGSPSPCSDRGVCMDSGSGQCLCRSGFAGTACELCAPGAFPGHCQAC 533

QY 878 SCSEHGQCDGITYSGECLCETGWTAAASCDPTTAVFAVCTPACSVHATCTNNTVCYNLN 937
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 534 RCTVHGRDEBLGSGSCFCDEGWTGPRCVEQLELPCTPPCAPEACRAAGNCECSLIG 593

QY 938 YESGIGTCTVVDPCQNNGGCAKVAKSGKGTQVSCSKKGYKDGYSCTEIDPCADGVN 997
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 594 YESGIRVCTVADLCDDGHHGCSSEHANCQVGTMTVCTCLPBYEDGWSCHARANPCTDHR 653

QY 998 GGCSEHATCRMTGPGKHKECKSHYVGVGDVDC-BPEQLPLDRCLQDNGOCHPDASCADLY 1056
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 654 GGCSEHANCSTGLNTRRCHECHAGYVGDGLQCLESESPVDRCLGPRPCHSDAMCTDLH 713

QY 1057 FQDTTGVFHLRSLPGLQYKLLTPDKAEACAKAEATATATYNQLSYAOKAKYHLCAGWLES 1116
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 714 FOEKRAGVFHLQATSPGGLNFSEBAEACEAGVILASFPOLSAQOGLFHILCLMGWLAN 773

QY 1117 GRVAVPTTYAOKGCANVYGVYDGSRAKNSSEMDVPCYRKADVNCCKAKAYVDDGS-C 1175
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 774 GSTAHPVVFPVADGNGRGVIVSGARKNISERMDAYCFRQDVACRCRNGFVGDGISTC 833
```

```

OY 1176 SGNLTQWMEPSTLNTNLEVLAFSSSSARGAFLKLTLDLSIRCTLFVPQNSGLPGKNS 1235
Db 834 NGKLLDVLATFANSTFYCMGLCYANATORGIDFLDFLDDLELYTKTLFVPVNEGFGVDNMT 893
OY 1236 LSGRDIIEHHLTNVNVSEFYNDLVNGTFLRMLMSQLITFFSODQLHQETRPVDG----RSI 1291
Db 894 LSGDEDELMHSMNTLLSAN-ASQGRKLLPAHSGSLIISAGCDNSMKAAPAGTIVVYSRI 952
OY 1292 LQMDITANGILHIISEPLRAPP-----TAATPAHSGLOTGTIFCAVVLVTGAIALAAVSY 1346
Db 953 TVMDIMAFNGIILHIALSPLLAPPOQAVLPAEPAPVAAGVGAVLAAAGALLGLLVAGALYLR 1012
OY 1347 FRLQRTTGFQRF 1359
Db 1013 ARGKPMGFGSAPF 1025

```

Search completed: June 17, 2002, 12:33:29
Job time: 300 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:28:54 ; Search time 29.51 Seconds
(without alignments)
4088.566 Million cell updates/sec

Title: US-09-842-930A-2
Perfect score: 7861
Sequence: 1 SLPSLLTRLEOMPDIYSIFRG.....WGHGPDMSQATTVTVPR 1431

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 256022 seqs, 84314167 residues

Total number of hits satisfying chosen parameters: 256022

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Pending_Patents_AA_New:*
1: /cgn2_6/ptodata/2/paa/US06_NEW.COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06_NEW.COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07_NEW.COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08_NEW.COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US09_NEW.COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US10_NEW.COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US60_NEW.COMB.pep:*

Pred. No. is the number of results predicted by the result to have a
score greater than or equal to the score of the change being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6072	77.2	1986	US-10-123-962-2	Sequence 2, Appli
2	6072	77.2	2017	US-10-123-962-4	Sequence 4, Appli
3	1094	13.9	510	US-10-104-047-2580	Sequence 2580, Ap
4	762.5	9.7	482	US-10-104-047-3794	Sequence 3794, Ap
5	480.5	6.1	1111	US-60-360-039-5280	Sequence 5280, Ap
6	399	5.1	3507	US-60-360-039-5784	Sequence 5784, Ap
7	375.5	4.8	1584	US-60-360-039-6549	Sequence 6549, Ap
8	373.5	4.8	1106	US-60-360-039-5709	Sequence 5709, Ap
9	371	4.7	3051	US-60-360-039-5866	Sequence 5866, Ap
10	359.5	4.6	1372	US-60-360-039-5971	Sequence 5971, Ap
11	338	4.3	3494	PCT-US02-09671-697	Sequence 697, App
12	338	4.3	3494	PCT-US02-09671-703	Sequence 703, App
13	338	4.3	3594	US-10-150-821-4	Sequence 4, Appli
14	338	4.3	3623	PCT-US02-09671-698	Sequence 698, App
15	338	4.3	3623	PCT-US02-09671-701	Sequence 701, App
16	338	4.3	3623	PCT-US02-09671-702	Sequence 702, App
17	335	4.3	3672	US-60-360-039-6146	Sequence 6146, App
18	334.5	4.3	3571	US-10-150-821-2	Sequence 2, Appli
19	333	4.2	1808	US-60-360-039-5986	Sequence 5986, App
20	331	4.2	3712	US-10-108-605-103	Sequence 103, App
21	326.5	4.2	1429	US-60-360-039-5819	Sequence 5819, App
22	326.5	4.2	3298	US-60-365-384-181	Sequence 181, App
23	326.5	4.2	3332	US-60-365-384-182	Sequence 182, App
24	325	4.1	2199	US-08-793-273C-2	Sequence 2, Appli
25	324.5	4.1	1810	US-08-793-273C-4	Sequence 4, Appli
26	324.5	4.1	2201	US-10-100-912-2	Sequence 2, Appli

27	322	4.1	1529	US-09-766-511B-29	Sequence 29, Appli
28	315.5	4.0	1247	PCT-US02-08253-216	Sequence 216, App
29	312.5	4.0	1512	US-09-970-944-32	Sequence 32, Appl
30	307	3.9	2823	US-60-360-039-5220	Sequence 5220, App
31	307	3.9	2823	US-60-360-039-5221	Sequence 5221, App
32	306	3.9	1529	US-09-312-283B-396	Sequence 396, App
33	299	3.8	988	US-60-369-876-2	Sequence 2, Appli
34	297	3.8	3871	US-60-360-039-6538	Sequence 6538, App
35	297	3.8	3871	US-60-360-039-6539	Sequence 6539, App
36	295.5	3.8	1534	US-09-970-944-30	Sequence 30, Appli
37	295	3.8	1295	US-60-360-039-5820	Sequence 5820, App
38	294	3.8	1295	US-60-360-039-5821	Sequence 5821, App
39	294	3.7	2715	US-10-042-865-52	Sequence 52, Appl
40	292	3.7	1531	US-09-970-944-31	Sequence 31, Appl
41	292	3.7	2824	US-60-360-039-5166	Sequence 5166, App
42	289.5	3.7	737	US-10-121-062-38	Sequence 38, Appli
43	288.5	3.7	1557	US-60-360-039-6816	Sequence 6816, App
44	286.5	3.6	838	US-60-360-039-6445	Sequence 6445, App
45	284.5	3.6	3501	US-10-123-155-37	Sequence 37, Appli

ALIGNMENTS

RESULT 1				
US-10-123-962-2				
Sequence 2, Application US/10123962				
GENERAL INFORMATION:				
APPLICANT: Fiddle, Carl Johan				
APPLICANT: Gerhardt, Brenda				
APPLICANT: Zambrowicz, Brian				
TITLE OF INVENTION: Novel Human EGF-Family Proteins and Polynucleotides Encoding t				
FILE REFERENCE: LEX-0338-USA				
CURRENT APPLICATION NUMBER: US/10/123,962				
PRIOR FILING DATE: 2002-04-16				
PRIOR APPLICATION NUMBER: US 60/286,141				
PRIOR FILING DATE: 2001-04-24				
NUMBER OF SEQ ID NOS: 5				
SOFTWARE: FastSeq for Windows Version 4.0				
SEQ ID NO 2				
LENGTH: 1986				
TYPE: PRT				
ORGANISM: homo sapiens				
US-10-123-962-2				
Query Match 77.2%; Score 6072; DB 6; Length 1986;				
Best Local Similarity 77.0%; Pred. No. 0;				
Matches 1083; Conservative 121; Mismatches 191; Indels 12; Gaps 5;				
QY	1	SLPSLLTRLEOMPDIYSIFRGYIHHNVLASATIESADAYTFVPNNEAIENYIREKKATSLK	60	
DB	571	SLPSLLTRLEOMPDIYSIFRGYIHHNVLASATIESADAYTFVPNNEAIENYIREKKATSLK	630	
QY	61	EDILRHVVHVLGKILKNDLHNGMHRETMLGFSYLLAFELRNQOLYNEAPIVNTVATDK	120	
DB	631	EDILRHVVHVLGKILKNDLHNGMHRETMLGFSYLLAFELRNQOLYNEAPIVNTVATDK	690	
QY	121	GVYHGLEKYLEIOTKNCNDNDTIIYRGEGCKSCSOAPCEPLETKPL-RETRKCIYSIFENG	179	
DB	691	GVYHGLEKYLEIOTKNCNDNDTIIYRGEGCKSCSOAPCEPLETKPL-RETRKCIYSIFENG	750	
QY	180	KKSVPFGCOPCVRITITRACHLASIAHNAKPAPEGVKMCALGTASVWDGVNGTGCQCQ	239	
DB	751	KRTLFPGCPKCVRIYVIRECCAGFEFGDPCPCPAAGVNCVCGNGICLDGVNGTGCCECG	810	
QY	240	LEFNGACETCEGKYGKIHCDQACSVHRCGSGPLGDSGCDQDVGWRCMDMETTNDN	299	
DB	811	BEFSGACTCEGKYGKIHCDQACSVHRCGSGPLGDSGCDQDVGWRCMDMETTNDN	870	
QY	300	CNGCTHSANCLLDPDGKASCKACAGFRGNGTVCTAINACETSNNGCSTKADCKRTTPCN	359	
DB	871	CNGCTHSANCLTNSDGTASCACAGFGNGTICTAINNCEISNGCCKADCKRTTPCR	930	

QY	360	RVCCKAGYGDGIVCIEIENPCLEHNHGGORNAECQOTGNOAVNCOLPKRYUDGKVCSL	413
Db	931	RVCCKAGYTDGIVCIEINPCLEHNHGGORNAECQOTGNOACNCLPAYTUDGVCCTL	990
QY	420	INVCITNNGGCSPPAFNCYTEEDODRITOCPRDYTGIDITVCRGSITGELPFPNSTSQVFO	479
Db	991	INVCITNNGGCSSEPAICNHTQVYRTOCTCKPNYIGDPSITGORSITGELPKNPNTSQVFO	105
QY	480	LOEHAVERLAGPGEFTVPAPLSSSENEHPRIKOWDOGIMSOYLRYHVGCQOOLLDNK	539
Db	1051	LOEHFVNDLVGPGEFTVPAPLSSAFDEARVKWMDKGLMPOYLRYHVACHOLLEENK	111
QY	540	VTSATITLQSGPVSISVSQDVFENNAKYLSDIISTNCVITHYDKLSPKULLTPMD	599
Db	1111	LISNATISQGEPIYISVSQDVFYNNNAKILSDIISTNCVITHYDKLSPKULLTPMD	117
QY	600	ALGRLOMLTVAAHNHGTFEFSKLIDQSGLSVTTDSIHPPVFMPTDKEALPPEOQ	659
Db	1171	NSGRLOMLTVLANNNGYIKFSNLIDQSGLSVTTDPIHPPVLFMTDQDALALAEQO	123
QY	660	DFLENDOKDKLSYLYKFHVIRDSKALASDLPRSASWKLQSGSELSYCTGSDIGELP	719
Db	1231	DFLENDOKDKLEKYEFHVIRDAKVALVLDPTSTAKMTQSGSELSYKCGAGHDIDLEL	129
QY	720	NEQMCRFIHRLELDVCAVAGSIDCLMLNPLTGGSCDFTTFEDIRBEGSGSIFPKPKLMS	779
Db	1291	NGQTCRIVORELLELDVCAVAGSIDCLMLDPTLGGSCDFTTFEDIRBEGSGSVNTPCPRMS	135
QY	780	KPKGVKKKCIYNPFPFRANNGCCNLTVYIQTFRCCGHGFMPDQCAPGPRTPCNRN	839
Db	1351	KPKGVKKCKCLN-LPFRKNLEGGCNERGSIVQIPROCKGTFGRNDQCAPGPRAPCNRN	140
QY	840	MCROLYTIPMGOCILCHTFGNJTACELCNHGRBGPDCQPRSCSEHGQCDDEGITSGECLCT	899
Db	1410	VCLDOYSATGECCKNTGFGNAGCEMCPWRFGRPCILPCGCSHDQOCDGITSGGQCLCT	146
QY	900	GMTAASCDTPAAVAVCPACSVNATCTENNTCVCNLYNMGDDGITCVVDFPCQONNGCA	959
Db	1470	GMTGPSCDTPQVAVPCTPPCSAHATCKENNTCECNLYDGDGITCVVDFPCQONNGCA	152
QY	960	KVANCOSQGTAVSCSCCKKGYGDDVSCIEIDPCADGVNNGCCHENATCRMTGPKHKCECK	1011
Db	1530	KVANCOSQGTAVSCSCQCKGYGDDHSHSTEIDPCADGLNNGCCHENATCKMTGPKHKCECK	158
QY	1020	SHVYGDVDCERBOLPDRCLQDNQCHCPDASCADLYFDOTYGVFHLRSPLQYKLTED	107
Db	1590	SHVYDGNCEBEPOLDRCLQDNQCHADAKCYDLHFODTYGVFHLRSPLQYKLTED	164
QY	1080	KAKKCAKEATITATPNYDLSYAKCAKXKHLCSAGLSESGRAVPTTYSQKCGANVGYTD	113
Db	1650	KAREKCAKEATATATPNYDLSYAKCAKXKHLCSAGMLEGRAVPAFPASQSCGGVGYTD	170
QY	1140	YGSRAKSEEMDVFECRYAKDVNCCCKAGYVODGSCGSLLOVLMSPSLNFTLEVLAF	119
Db	1710	YGPBRKSEEMDVFECRYAKDVNCCCKAGYVODGSCGSLLOVLMSPSLNFTLEVLAY	176
QY	1200	SKSSARGAFLTKHLTDLISIRGTLEVPONSGLPNGKSLSGNDIEHHLTVNVSEYNDLVNG	125
Db	1770	SNSSARGAFLTEHLLTDLISIRGTLEVPONSGLENETLSGRDIEHHLTVNMFYNDLVNG	182
QY	1260	TFLKRLMCSOLLITFSQDLH-QETRYVDGSLILOMILIAANGLIHISEPRLAPPAAT	131
Db	1830	TTLQTRLSKSLITLIASODRPTLETRFVGDRAIILOMIPISANGLIHISRLAPPAVIT	188
QY	1319	AAHSGLSGIGICAVLYVGAIALAAYSSEFRKORTGTGFORPDOKRTLMWLLAS-----	137
Db	1890	LTHSLGAGITFAIILTVGAALAYSTFRNKRRTIGFOHFESEEDITNVALGKOEPENT	194
QY	1373	SSPRTISQTLCKMRPQRRHPOSPVPTPSQ	1399
Db	1950	SNPLTESITTSAPPE---PSYDPTDSE	1973

```

RESULT 2
US-10-123-962-4
; Sequence 4, Application US/10123962
; GENERAL INFORMATION:
; APPLICANT: Friddele, Carl Johan
; APPLICANT: Gerhardt, Brenda
; APPLICANT: Zambrowicz, Brian
; TITLE OR INVENTION: Novel Human EGF-Family Proteins and Polynucleotides Encoding t
; FILE REFERENCE: LEX-0338-USA
; CURRENT FILING DATE: 2002-04-16
; PRIOR FILING DATE: 2001-04-24
; NUMBER OF SEQ. ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2017
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-123-962-4

Query Match 77.2%, Score 6072; DB 6; Length 2017;
Best Local Similarity 77.0%; Pred. No. 0;
Matches 1083; Conservative 121; Mismatches 191; Indels 12; Gaps 5;

QY 1 SLPSLLTLEDMQDPVSTIRGYIIHNFLASIESADYVFPNNPAIENYIREKKATSLK 60
DB 602 SLPSLLTLEDMQDPVSTIRGYIIQYINLALAEADYVFPANNALAIENYIREKKVLSLE 661
QY 61 EDILYVHVVLGEKLLKNDLHNGMHRETMLGFSYLLAFPLANDOLYVNEAPINTYVATDK 120
DB 662 EDVLKHYVLEKLLKNDLHNGMHRETMLGFYSYLLFPLANDOLYVNEAPINTYVATDK 721
QY 121 GYHGLEKVLLEIQNRCDNNDTIYRGECKGCSQAQAPLETKPL-RETRECIYSIRPMG 179
DB 722 GYHGLEKVLLEIQNRCDNNDTIIRGRCRICSSLETCPFGTKSLGNEKRRCIYTSYPMG 781
QY 180 KRSVITGGOPOCVRIITIRACMLASLANHAPAPPEVVMCALGASVMDGVNGTGCCQG 239
DB 782 RRTLTIGOPKVPKRVITIRRECCAGFFGPOCOPCGMAONVCFGNGICLDGYNGVGCYCECG 841
QY 240 LGFNLTACETCTEGKYGIHCDQACSCVHGRCSSQPLGDSGCDVDYGMWGVKCDMEITTDN 299
DB 842 EEFSESTACETCTEGKYGIHCDQACSCVHGRCNGPLGDSGCDVDYGMWGVKCDMEITTDN 901
QY 300 CNGTCHTSANCLLDPDGKASCKCAAGFRNGTIVCTAIVACETSNGGCSTKADCKRTTGN 359
DB 902 CNGTCHTSANCLTNSDGTASCKCAAGFQNGTICTAIVACETSNGGCSTKADCKRTTGN 961
QY 360 RVCVCKAGYTGIGVYCLEINPCLLENHGGCDNRNAECTOTGPNDAVANCCLPKTNGDKVCSL 419
DB 962 RVCVCKAGYTGIGVYCLEINPCLLENHGGCDNRNAECTOTGPNDAVANCCLPKTNGDKVCSL 1021
QY 420 INVCLLTNNGGSPFAFCNYTTEODORICTCKPDYTGIDGIVCRGSIYELDKPKNPSTSQYEFQ 479
DB 1022 INVCLLTNNGGSEFAICNHTGTGVEKTCCKPNIYIDGFTCGRSIYQELDKPKNPSTSQYEFQ 1081
QY 480 LOENHAVRELAGPRTTVAPRLSSSFNHEPRIKMDQOGLMSOVLRYHYVGGCOQLLDNMLK 539
DB 1082 LOENHAVRELAVGPPRTTVAPRLSAAFDEARVYKMDQVGLMSOVLRYHYVAAOHDLLEMLK 1141
QY 540 VTTSAFTTLOGRPVSTISVSODPVEFINNEAKVYSSDIISTNGVHYVDKLSPRKNLLITPKD 599
DB 1142 LLSNATSIQGEETIVSVSOSTIYVINKKAKIISSDIISTNGVHYVDKLSPRKNLLITPKD 1201
QY 600 ALGRVLIQNTLVAAHNGHYTKFSKLLIQDSGLSLVITDSIHTPTVYFWPTDKALALPPEEQ 659
DB 1202 NSGRLLQNTLVTAHNGHYTKFSNLLIQDSGLSLVITDPIHTPTVLEFWPTDQALHALPAPQO 1261
QY 660 DFLNODNKKDKIKSLYKTHVITRDSKALASDLPRKASMKMTLOGSELSVYCGTSGSDIGELFL 719

```



```
Db 1262 DFLFNODNKKDKLEKYLKFEHVIIRDAKVLAVLDLPTSTAMKTLQSGSELSVKCGAGRIDGLFTL 1321
Qy 720 NEOMCRFTHRGILFDPVAVAYIDCLLMNPDTGGRCDDTFETPDLPBEGSGCIFTPRCPLKS 779
Db 1322 NGQOTRIVQRELLFBDLVAVYIDCLLIDPFTLGRCDDFTFTFASGECGCVNTPSCPRMS 1381
Qy 780 KPKGKKKCIYVPLFRFNNVEGCONLCTVVIOTPRCCHGFMFDPDQACGGGPDTCNNNG 839
Db 1382 KPKGKKKCLYN-LPFKRNLECCRCRCSLYIOTPRCKCYFERDDQACGGGADACNNNG 1440
Qy 840 MCRDLTPYMGOCCLHTGNGTACELCMHGRFGPDCQPRSCSEHGQCEGITSGECLCET 899
Db 1441 VCLDQYSATGECCKNTGNGTACELCMHGRFGPDCQPRSCSEHGQCEGITSGECLCET 1500
Qy 900 GMTASCTPPAVAFVAVCPACSVHATCTENNTVCNLTNEGITCTVVDPECKONNGCA 959
Db 1501 GMTGSCDQAVLPAVCGTPPCSAHATCKENNTCECNLDL EGGQITCTVVDPECKONNGCA 1560
Qy 960 KYAKCSOKGTQVSCSKGKGYKGDYSCIEIDPCADGVNCGCHEHATCRMTGPKHKCECK 1019
Db 1561 KYARCSOKGTQVSCSKGKGYKGDYSCIEIDPCADGVNCGCHEHATCKMTGPKHKCECK 1620
Qy 1020 SHYVGDVGDCEBEQOLPLDRCLQDNGQCHPDASCADLYFQDTTVGVFHLRSPLGQYKLPFD 1079
Db 1621 SHYVGDGNCBEQOLPLDRCLQDNGQCHADACVDLHFQDTTVGVFHLRSPLGQYKLPFD 1680
Qy 1080 KAKECACAEATATVYNOLSVOKAKYHLCISAGMLESGRVAVPTTYASCKGCANVGYVD 1139
Db 1681 KARECACAEATAVYNOLSTAOAKAKYHLCISAGMLETGVAVPTTYASCKGCANVGYVD 1740
Qy 1140 YGSRAKSEMDVFCYRKMDVNCCTKAGYVGDGFCSGNLLQVLMSPSLTNFLTEVLAF 1199
Db 1741 YGPRNKSEMDVFCYRKMDVNCCTKAGYVGDGFCSGNLLQVLMSPSLTNFLTEVLAF 1800
Qy 1200 SKSSARQGAFLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNG 1259
Db 1801 SNSSARQGAFLKHLTDLSTRTGLFVPONSGLGENETLSGRDIEHHLTNVNSFYNDLVNG 1860
Qy 1260 TFLPMLCSOLLITPFSOOLH-OETRFVDSRSILQMDITIAANGILHISEPRLAPPTAT 1318
Db 1861 TTLQTRLSKLLITASODPLOPTETRFVDGRAILQMDITFASKGILHIVSRPLKAPAPVT 1920
Qy 1319 AAHSGLTGIFCAVVLVTGATALAAVSYFRLKQRTTGFQRPDOKRTLMWMLAS----- 1372
Db 1921 LHTGLGAGIFAILLVGAVALAASVFRINRRTIGFQHPSEHEDINVAALGKQOQPENI 1980
Qy 1373 SSPRISQTLCKMRPQRHROSPPVPSQ 1399
Db 1981 SNPLYESTTSAPPE--PSYDPTDSE 2004
```

```
RESULT 3
US-10-104-047-2580
; Sequence 2580, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2580
; LENGTH: 510
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-2580
```

Query Match 13.9%; Score 1094; DB 6; Length 510;
Best Local Similarity 47.8%; Pred. No. 7.7e-54;

```
Matches 193; Conservative 60; Mismatches 149; Indels 2; Gaps 2;
Qy 849 GGCCLHTGNGTACELCMHGRFGPDCQPRSCSEHGQCEGITSGBRCLEETGWTASCDPT 908
Db 8 GGCCLHSCFGAGTACELCAGAFGPHCQACRVTYHGRCDLGGSGCFDEGCTGRCEV 67
Qy 909 PTAFAVCPACSVHATCTENNTVCNLTNEGIDTCTVVDPECKONNGCAKAVAKCSOKG 968
Db 909 PTAFAVCPACSVHATCTENNTVCNLTNEGIDTCTVVDPECKONNGCAKAVAKCSOKG 968
Qy 969 TQVSCSKKGYKGDYSCIEIDPCADGVNCGCHEHATCRMTGPKHKCECKSHYVGDVD 1028
Db 128 TWVTCTCLPBYEGDWSGRARNPCTDGHGSGSEHANCSTGINTPRCEHAGYVGDGLQ 187
Qy 1029 C-EPEQLPLDRCLQDNGQCHPDASCADLYFQDTTVGVFHLRSPLGQYKLPDCAKACAK 1087
Db 188 CLEESPPVDRCLQDPPRPHSDAVCTDLHFQERKRAVFNHQAATSGYGGINFSEAEACAA 247
Qy 1088 EATATVYNOLSVOKAKYHLCISAGMLESGRVAVPTTYASOKGCANVGYVDYGRANKS 1147
Db 248 QCAVALASFPQLSAADQLGFHLCLMGLANGSTAHVVFVADCGNGRGVVISGARKNIS 307
Qy 1148 EMDVFCYRKMDVNCCTKAGYVGDGFS-CSGNLLQVLMSPSLTNFLTEVLAFSSKASRG 1206
Db 308 EMDVFCYRKMDVNCCTKAGYVGDGISTCGKLLDLVLAATANFSTFGMLGVAANTORG 367
Qy 1207 QAFKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNG 1250
Db 368 LQFLDLDELTYKTLFVPONSGFVDMNPLSGDLELHASNATL 411
```

```
RESULT 4
US-10-104-047-3794
; Sequence 3794, Application US/10104047
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: Novel full length cDNA
; FILE REFERENCE: H1-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
; PRIOR FILING DATE: 2002-03-25
; CURRENT APPLICATION NUMBER:
; PRIOR APPLICATION NUMBER:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3794
; LENGTH: 482
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-104-047-3794
```

Query Match 9.7%; Score 762.5; DB 6; Length 482;
Best Local Similarity 38.6%; Pred. No. 2.3e-35;
Matches 164; Conservative 57; Mismatches 159; Indels 45; Gaps 8;

```
Qy 971 VQSCCKKGYKGDYSCIEIDPCADGVNCGCHEHATCRMTGPKHKCECKSHYVGDVDC- 1029
Db 2 VTCTCLPBYEGDWSGRARNPCTDGHGSGSEHANCSTGINTPRCEHAGYVGDGLQCL 61
Qy 1030 EPEQLPLDRCLQDNGQCHPDASCADLYFQDTTVGVFHLRSPLGQYKLPDCAKACAKEA 1089
Db 62 ESEPPVDRCLQDPPRPHSDAMCTDLHFQERKRAVFNHQAATSGYGGINFSEAEACAA 121
Qy 1090 ATATVYNOLSVOKAKYHLCISAGMLESGRVAVPTTYASCKGCANVGYVDYGRANKSEM 1149
Db 122 AVLASEPQLSAADQLGFHLCLMGLANGSTAHVVFVADCGGRGVVISGARKNLSER 181
Qy 1150 WDFCYRKMDVNCCTKAGYVGDGFS-CSGNLLQVLMSPSLTNFLTEVLAFSSKASARGQA 1208
Db 182 WDAVCFRVDVACRCNGVYGDGISTCGKLLDLVLAATANFSTFGMLGVAANTORG 241
Qy 1209 FLKHLTDLSTRTGLFVPONSGLPQNKSLSGRDIEHHLTNVNSFYNDLVNGTFLRTMIGS 1268
```

Db 242 FLDELDELTYKTLFVFNVEGVNDMTLSGPDSELSHASNALLSAN-ASQKLLPAHSGL 300
QY 1269 QLLIT-----FSODQLHOETRFYDGRSILQMDIITA 1298
Db 301 SLIISDADPDNSMAPVSLATVPFCWMPSPSLTVGLTLVSLGAPGVYSR-IYWDIMA 359
QY 1299 ANGLIHIISEPIRAPP-----TAATPAHSGLCTGIFCAVAVLTGAI--ALAAVSFRLKQ 1351
Db 360 FNGIHAALASPLAPPOQAVLAPAPVAAAGVG---AVLAAGALLGLVAGALYLRAG 415
QY 1352 RTTGF 1356
Db 416 KPTGF 420
RESULT 5
US-60-360-039-5280
; Sequence 5280, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5280
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5280
Query Match 6.1%; Score 480.5; DB 7; Length 1111;
Best Local Similarity 22.2%; Pred. No. 3,1e-19;
Matches 251; Conservative 98; Mismatches 341; Indels 441; Gaps 70;
QY 115 NWATDKGVY-HGLEKYLEI-----QKNRCN-----NDTIYRGECKGSGQAPCLETRK 163
Db 43 HVCITVKTIVDDYELKKVHIHTVYNDTEQCLNPLTGFQCTVERKQ--KASYQRQL--VK 97
QY 164 PLRETRKCIYIYFMGRKRSVFIGCOPGV--RTIIRACWLASLANAPAGEVMCA- 220
Db 98 KERIVKQCCDGYOTKHCLPDCNPPCKKKKCIERKQ-----ECDPYIGG-KYCAS 149
QY 221 -----LGTASVWDVNGT-----GTCCGGLGFNTACE-TCTEGKYGIHCDQAGSC 265
Db 150 SCVSQGTWGLGCKSCSCCENGANDPELGTICITSGFGGERCEPCPDNMGWPCVSKSCPC 209
QY 266 VH-GRCSQPLDGDSCDCDVGR-----GVYCDMEITITDNCNGCHTSANCL 311
Db 210 QWNGCKNK-----EGKVCSDGWDGGEFCCLKCEBKFGAECKPE-----CN-CQNGATC- 257
QY 312 LDPDGKASCK-----CAAGFRGNGTV--CTAIN--ACETISNG----- 344
Db 258 DMTNGKCIKSGYHGLACENECVSFGFGSGCTQKCDCLNNQNDSSGECKCICTGWTGKHC 317
QY 345 --GCST--KADCKR--TTPG-----NRVCYKAGYTGDIYCLEINPLEN 384
Db 318 DIGCSHGRFGLQCKQKCTPGLEFSDSNASCDAKTGCCQCESGKGP--KCDERKCDARQ 375
QY 385 HGG-CDNAECYD-----TGPNAVCNCLPKYTGDKGVSLINVCILTNNG--CSPFAPC 436
Db 376 YGADCKCTCTCAVRENTLMCAPNTPGFCRCKPGEYGDN--CEL--ACSKDSYGPCEKQIAC 431
QY 437 NTEDOQ-----RICTCKPDYTGDIYCGSISYGLPKNPSTSOYFQLOEHAVRELACG 491
Db 432 DMNHASECNPETGSCVCKRFGRTG-----KNCSE-----P 460

QY 492 GPFTVFAPLSSSFNHEPRKMDDOGLMSOVLRYHWGCOQLLDNKLKATTSATTLQGE 551
Db 461 CPLEDYGP--NCAHQCC--NORC-----VGD----- 484
QY 552 VSISSQDTVFINNBAKVLSDIISTNGYIIVIDLKLSPKNLLITPKDALGRVLONLTV 611
Db 485 -----GADCKQCCDGMWGH 499
QY 612 ANNHGYTRKSLIODSGLSVITDSIHPPVTFWPTDKALEALPEQODFLPNODNKL 671
Db 500 RCEH-----HCPADTF----- 510
QY 672 KSYLKFHYIRDSKALASDLPRASWKTLOGSELSVRCYSGSDIG--ELPLNQMCRFIRNG 730
Db 511 -----GANCKRCKCPKGIQCDPTTGECTCPAGLQ 541
QY 731 LLEFVGV--AYGIDCLMNPILGRCDFTFTFDIPGEGSC---IFTPKPKLK-SKPKG 783
Db 542 ANCDIGCEPGSYGPGCKLHCKCYNGKCKQET-----GEC-TQOPGPFSSDCSTTSKGR- 594
QY 784 YKKKCIYNELPFRNVEGQNLG-----TVIYQTPRCCHGYMPDQACPGGP-----DT 833
Db 595 -----YG-----ESCELSGCSGSDASCSKQTKC-----LCPLTKGVSCDQ 630
QY 834 PC--NNRG-MCRDLYTPM-----GCCL-CHTFGNFACE-LQWHRFGPDOP-RS 878
Db 631 KCDPNTFGLCQETVTPSPCASTDPKNGVCLSPRESSGHEHNCPASYSDDGCOQVCS 690
QY 879 CSEHGQCDGRTGSGECLEGTMTAASCD--TPTAVF-----AVCTPACSVHATCTE-NTT 931
Db 691 CADGHCD---PTTGECICEPBGYHGKTCSEKCPDGKYGIGCALDCKKCKAGSGSTCHINL 747
QY 932 CVCNLTNEBGDITGVVDECKONNGCAKAVAKCSOKGTVS-----CSCKKYGKDGYSK 986
Db 748 CIOPAGE--GALCTRPCSAGFMNGCRQVCRCTSEYKQCMQOTGBCSPAGFOGD--RC 803
QY 987 IETDPCADGVN-GCHEATCKMTGPG-----KHKCEKSHYVD-----GVNC 1029
Db 804 DK--PCDEYGPDCIKKCKCGGTATSSCNRVSGACHCHPGFTGECHALCESTFFGLKC 861
QY 1030 EPEQLPLDRCLDNGCHPDASCADLYFODTIVGVFHL-RSPLOYKLTFLD 1079
Db 862 SKF-----C-PKDCGSGDYECDDAIGCHVDMCGKRAKQFE 898
RESULT 6
US-60-360-039-5784
; Sequence 5784, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5784
; LENGTH: 3507
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5784
Query Match 5.1%; Score 399; DB 7; Length 3507;
Best Local Similarity 16.7%; Pred. No. 3,8e-14;
Matches 316; Conservative 153; Mismatches 490; Indels 932; Gaps 74;
QY 134 KNRCDNNDTIYVGECKGCSQAPCLET-KPLRETRKCIYIYFMGRKSVFIGQCP--Q 190

```

Db 1537 KKPDSTOS--SKSHCESNMS--CEVDTVDGSECKEOMGGYKKSGK-----VCEDINE 1587
QY 191 CVRFTITRACMLASLAHNAKPAPGEVKKMCA LGTASVMDGVNGTGGCCGLGFNG----- 244
Db 1588 CV-----AEKAP-----CSLANCV--NMNGTSCSCOGYRGDGFMCFT 1624
QY 245 --TACE-----TCF--EKKYGIHC-----DOAC----- 263
Db 1625 DINEDEHHPCHPAHECTNLBESSFKECHSGFEGDCKIKCTNPLERSCDEYKFCGRVDH 1684
QY 264 --SCVHGRCSCQPLGDGSCDDCVGMRGV--CDMEITDNC--NGCTHTSANCLLDPP 315
Db 1685 VCSLVRIRYNGSL--SSVCECEGFPEKESNSC--VIDECEBSRNNCDPASAVCVTE 1740
QY 316 GKASCKAAGFRNGCTVCTAINACTSNGGCTKADCKRTTGTGNVAVCKACAGTGDGIVC 375
Db 1741 GSYRCECAEGTEGEGVCTDIDECRGAGCDSMACINRM--GSCCCKCMAGYTDGATC 1799
QY 376 LEI-----NPLENH----- 385
Db 1800 IKIEEPKSDKTACTDEMSRLCELEKKOCTVDEEEVPOCGACLPGHHPINGTCSQIOTSG 1859
QY 386 ----GGCDRNAECTQTGPNOAVNCCLPKRYTGDGVCSLINVCLTNNGCCSPAFACN--Y 438
Db 1860 LCAOKNDCKNHAECIDIPDISHFCSPDGFIDGMI CDVDEC--NNAG-----MCDDEN 1912
QY 439 TEODORI--CTC-----KPDYTC- 454
Db 1913 TKCEMTIOSFNCVCLLEGKKYDEKCVDEKKOPNREKIEIDEENSSSSNGOEKPTTKGI 1972
QY 455 ----- 454
Db 1973 VSSTSATSESTTAPHTTSSITTTKDMTSSKSPENVMTSSSPREYTSSTSSKSTTAS 2032
QY 455 -----DGIVCRGSIYELPKPNSTQVFPOLQ-----E 482
Db 2033 ETTVSSPSESSSEAPLTSSPATTEVTETSSVSKSTPKESSESEITVKLSSKPEYTE 2092
QY 483 HAVRE-----LAGPFPYFAP-----LSSSFN- 505
Db 2093 SSVASSPSTTSSQSVTSTVPETSKSTVLSSEAPVSTSPTEVHTSSETPKPSLASSST 2152
QY 506 -----HEPRIKDM-----OOG 517
Db 2153 GDTNSTPTSSLASVSKSTSAPEGTSASAVAPKYLSSLPDVSOPTKTFDATESSTVQAS 2212
QY 518 LMSOVLRIHVHVCQOOLLNDLKVTT-----ATTLOGEPV 552
Db 2213 ETSSGTSVSKSTSEPSHVTKLSTSSNPSSSVPTSPKSTPTVPESTEQPTSTPSSQSL 2272
QY 553 SISVSODTVFINNEKAVLSS----- 572
Db 2273 TPMSNSSEVLTSTSEPHVLSSSLSPDVSSOSTTPNNLSESTVETPKTSEVSLNSEEST 2332
QY 573 -----DIIS-----TNGVIVHIDKL 588
Db 2333 TEAPPTLSPLDLSSTNNLSOSSYSTEDRSRISSENSEKPTSAPELVTSSVTHVAASSP 2392
QY 589 SPKNULITPKALGRVLONL-----TTVAANH----- 615
Db 2393 DVPTSESEPDLTGSGTENIPEASSKOTISSTPPDTTTASSEPKSTSMSPDLSTSNV 2452
QY 616 -----GYTKFSKLQDQGLSVITDSI--HTPVVFMPT 647
Db 2453 LSESTTPSESSKSPVSSSTBGISVYSTEFSK--PESTISSVLEEDLTKTTPSPILEET 2511
QY 648 DKALEALPPEODFL----- 662
Db 2512 TTASPTSPLEEDSLTVSVRIHELTTSENVPKSESTTTSSSESSKSPQEPAGILITTV 2571
QY 663 -----FNODN-----KDKLSYIKFHVIR----- 681
Db 2572 VPTSSVSLITASEIEAITSNTPFKOGRPTTTPSKSLVKSSTTSPSTVSSSESTKRTT 2631

```

```

QY 682 ----- 681
Db 2632 VSTTVSTTTPTTETTTSELIIITAAPSKFTTESTESSEAPPTTPATSETKPSNVSTSK 2691
QY 682 -----DSKALASDLPKRSAMKTLGSELS 705
Db 2692 STENETSTNSQGSLESTMSSTSEPETNAPAVTVSSASSTTLEPNSTSSPTSSSEAS 2751
QY 706 VACGSGSDIGELFLNE-----OMCRFIRGLLF-----DV 735
Db 2752 YK--LSSLFPESITSEAVTVSSRAPELITMSSESHREISTVSSSEPEEIPLTSTVSPV 2809
QY 736 GVAYGI-----DCLLMN--PTLGRDQFTTFDIPGEG 767
Db 2810 VTASSIPSEEPILLSSVTSSSTPRVRLTGTPODLIVSTVPSHGNNRONTASSVPSNST 2869
QY 768 SCIF-----TPKCP--LKSXPKGVKK-----CIYNPLP-----FR 796
Db 2870 SPILPSESLTTPQPPPTTTAKPATTSCKRGPPSIQPPAEMFTTPAPPPSNGCYGE 2929
QY 797 RNEBCCQNLCTVVIOTPRCHGYFMPDQACPGPDTPCNNRGMKRDLYTPMGQCLHTG 856
Db 2930 TNOEBEQVSTTTTEAPSLC-----STVYCHSLATGEQ--STGVICICRDG 2972
QY 857 ENG--TACELCWHGRFGPDC--OPRSCSEHGCGDEGITGSECLCETGWTASCDTPTA 911
Db 2973 FTGDGTTMCS--KSTADCLSLPSLCADKACD--NSTSCBDDAGYIGDG----- 3019
QY 912 VFACVP-----ACSVHATCTENNTVCNLYEBGDIITCVVVPFCOKONNGCAKV 961
Db 3020 --YVCSPPHQDCVLKNDLNSPEAVC-QNRROCLPGFTGDDGVKCVSIHERASNGQCDAN 3076
QY 962 AKCSQKGTQVSCSKKKGKGDGYSCTEIDP--CADGVNGGHEHATCRMTGPKKCKCK 1019
Db 3077 AHC-VGGT--TCKCNPGYFGNLCCVP--DPLDCVH-FTGICHPNAYCNSE--SHQCCS 3128
QY 1020 SHYVGDDVDCPEPEOLPDLRCLQDNGCQHPDASCADLYPODTVGVFHLRSPIGQ----- 1073
Db 3129 SEFSGNGVSCFPQK--SCRTDKSVCAKMAICL-----PQSCICIRMG 3168
QY 1074 YK-LTFEAKAKEACAEATTIATYNOLSYAQ--KAKYHLCASG--WLESGRVAYPTTYAS 1127
Db 3169 FKGDPPFYKCSLVSKEPANQODLSVSSCVTPPCDSDSTQICISGDECICKSGFRNRNTLSS 3228
QY 1128 QKCGANVVGIVDYSBRANKSEMMDVFCYRMKDVN-----CTKAGYVGDGFCSCGNL 1180
Db 3229 ETC-----ADIDCAEAKSHK-----CDRVATCRNTFGSHVCTCPDGHVGDGITCVPHVN 3277
QY 1181 QVLMSEFSLTNFLTEVLAFSKSSKASGOAFLK 1211
Db 3278 QGKLSVYCEPADGMDTLVIGNETADFEKILFVK 3308

RESULT
7
US-60-039-6549
: Sequence 6549, Application US/60360039
: GENERAL INFORMATION:
: APPLICANT: Cao, Yongwei
: APPLICANT: Chen, Xianfeng
: APPLICANT: Goldman, Barry S.
: APPLICANT: Hinkle, Gregory J.
: APPLICANT: Slater, Steven C.
: TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
: FILE REFERENCE: 38-10(52052)A
: CURRENT APPLICATION NUMBER: US/60/360,039
: CURRENT FILING DATE: 2002-02-21
: NUMBER OF SEQ ID NOS: 47374
: SEQ ID NO 6549
: LENGTH: 1584
: TYPE: PRT
: ORGANISM: Caenorhabditis elegans

```

US-60-360-039-6549

Query Match 4.8%; Score 375.5; DB 7; Length 1584;
 Best Local Similarity 20.2%; Pred. No. 3.3e-13;
 Matches 242; Conservative 125; Mismatches 384; Indels 449; Gaps 65;

18 FRGYIIHYNLASAIESADAYTVFVPPNEAI-----ENVIREKK---ATSLKEDIIIRYHYV 69
 502 FSGRNIDHFYKSHLFVSGTLPIAPGSEVIIPDEEVEYRERRGVLTSAARDVI----- 556
 70 IGEKLLKNDLHNGMREYMLGFSYLLAFLLRNDQLYVNAPIYNTYNVATDKGYHGLEVY 129
 557 -----VRDGNVYOT-----YRMSV---DQOITFEBCP---NKEPDRD--HSMK-- 591
 130 LEIOK-NRCNDNDTIIIRGECCGCSQAAPCLETKPL--RETRKCIYSIYFMGKRKSVFTG 186
 592 LHWKRLNYYVNDDEGVYRKGAKNFATRSVGAIVASAPSGHFPRROHOGSNMASERPIETP 651
 187 COPQCVRIITTRACWLASIAHNAKPAPEVVKCALGTASVMDGVNGTGCOCGLGFN--- 243
 652 SQSESIST-----DSVCAPGR-HQCTLPNMKC-RVYDPSYRCCEPEGYQAHH 696
 244 -----GRACETCTGKYGIV-CDQACSCVHGRCSGGFLGDSGCDGVWR--GYKDMH 294
 697 DASSHIGWICODLDECOGDHNCDO-----HAKCTNRP-GSFSCQCLQGYGDGRSCIRE 750
 295 -----ITTD-----NCGTCHTSANCLL--DPDGKASCKCAAGFRG 328
 751 HOASHHNNQOTPOEMAGYGATTEGECTAHNOCHEMGECVFTSEHPTGR--CKCKRWYWG 808
 329 NGTVCTAINACE-----TSNGG-----CSTKADCKRTTPGNRYCVYCKAGY 368
 809 DG-----VNHGCRPRENNMPKHANIPQRGQAGSYVCVNAECMPEPSGSECVYCKAGF 863
 369 TGDGIVCLEINCLNHHGGCDRNAECTOTGPNQAVCNCLPKTTGDKGKVSLLINVLITNNG 428
 864 SGNVGYTCESLDDRHAHSHSNHEDQOQGT-----SLGKYCRSHDEC-SEHG 909
 429 GCSPAFCNTEPDORICTCKPDTYGDGIVCRGSIYGLPKPNSTSQYFQLOEHAVREL 488
 910 SCAYHSHSLGYO-----CACTEPIYVNGNYEC-----TLPSSAS----- 943
 489 AGRPPTVPAPLSSSFNHEPRIKMDQGLMSQVLRHVHVGCOQLLDLNLKVTTSATTLQ 548
 944 -----AP-----ELGCD-----YTRDC----- 955
 549 GEPVSIYSODIVFINNEAKVLSLDSISTNGYIHYIDKLSPKNLLITPKDALGRVQLNL 608
 956 -----SEPADCVYERS-----STGATFRCV----- 975
 609 TVVANHGYTKFSKLIQDSGL-LSVITDSIHPTVYFMPTDKALEALPEEODFLFNQDN 667
 976 -----QOSGYTGDKGYCKMQSOLAISAIAPSV-----PQLPSEP----- 1008
 668 KKKLKSYYLKFHVIKRSKALASDLPRSASWKVLQGSFLSYRCSTGSDIGELFLINEMCRFI 727
 1009 -----AVLSTASCNFCNCPDAQCVDYDHNRC----- 1034
 728 HRGLLFDVGVAYGIDCL-LMNPITLGRCDTFTTFDLPBGSCGCTFPKPKLAKSKPRGYVK 786
 1035 YNCECYAAMPDGMGNCVPLAKPNM-----VPAQPKTCVSSDCHING----- 1076
 787 KCIYPLPFRFRRNVEGCQNLCTVVIOTPRCHGYFMPDQACAG-----GP 831
 1077 HCVINE-----HGAGEYICQCLGFSGDSGFINCRGADQCNP 1112
 832 DTP--CNNRMC-RDIYTPMGOCCLHTGF--NGTACELCWHGRFGP---DCQPRSCSH 882
 1113 SNPSACYQNAHCYUAILNAHACKCVDGFKGDTGTCV---PYAPATNCNLEPRICHAN 1167
 883 GGC-----DEGITGSGECCLCFETGWTAA--ASCD---TPTAVFAVCTPACGVHATCTENNT- 931

DB 1168 AOCVMHND--TNAYECLICKPGSSGDGYTKCDVIETPR-----CT-NCSTIAHYCAQNPPTS 1218
 QY 932 -----CWCNINTEGEGITCTVVDPECKKONGGCAKVAKC--SOKGTQVSCCKKGYKGDGYS 985
 DB 1219 GAYQCKCMANGYNGNGLCYVSMSSCLDDRLSDENMADCVGEAGHY-CNCHGYHDDGRS 1277
 QY 986 CI-EIDPCADGVNCGGCEHATCRMTGP--GKNKCECKSHY-VGDGVDCPEPQLPLD--- 1037
 DB 1278 CSPESSTRSDKLLVARGMAIFERSTNPDEYKQLIVIPHHIPVGLDFDCKEEKIWSDMS 1337
 QY 1038 -----RCLDNGQCHP-----DASCADLYEODT---TVGYFHLRSLPGGYK 1075
 DB 1338 GHSITRSSLNGTEHMSYFNKEKLSPEGLAVDWSRRVYADSNDELGVASLN---GKPK 1394

RESULT 8

US-60-360-039-5709
 Sequence 5709, Application US/60360039

GENERAL INFORMATION:
 APPLICANT: Cao, Yongwei
 APPLICANT: Chen, Xianfeng
 APPLICANT: Goldman, Barry S.
 APPLICANT: Hinkle, Gregory J.
 APPLICANT: Slater, Steven C.
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 FILE REFERENCE: 38-10(52052)A
 CURRENT APPLICATION NUMBER: US/60/360,039
 CURRENT FILING DATE: 2002-02-21
 NUMBER OF SEQ ID NOS: 47374

SEQ ID NO 5709
 LENGTH: 1106
 TYPE: PRT
 ORGANISM: Caenorhabditis elegans
 US-60-360-039-5709

Query Match 4.8%; Score 373.5; DB 7; Length 1106;
 Best Local Similarity 19.3%; Pred. No. 2.9e-13;
 Matches 193; Conservative 85; Mismatches 265; Indels 459; Gaps 42;

230 VNGTGCOCGLGFNGTAC-ETCTEGKYGIHCDAQ-----SCVHGRCGSGPL 275
 173 VGDRIYLCGAGFKGLECIESCAPSQWFNDCKSKCHDGDGSDCDEBNGRCPNKCSPGWI 232
 276 GDGSCDCDVGMRGVKCDMEITTDNC-----CKCAAGFRGNGTGTATN 337
 233 GEPICDEMD-----ECEMGI--DNCPRNBPDLNTPGSLCLFEYDEAQQCKNSKAP 286
 301 -----NCTCHTS----- 307
 287 PSAPLIPVDVIPMHPPTPKPTPPKYVAPPRNRFKSTTAATSPFITETTRTOTTTREKS 346
 308 -----ANCLLDPOCKAS-----CKCAAGFRGNGTGTATN 337
 347 ISTPLITTEVTVRPSEPTTVVTAAPACARC--DQNAKCSNGVCTSGEFTTGDFRCYDAD 404
 338 ACE-----TSNGG-----CSTKADCKR----- 354
 405 ECEIIRAVCRQHSICSNTIGSFECTCHGYRREDGKCEVDECRLEPKICGDPNKTGKI 464
 355 TTPGNRVCYKAGYTG-DIVCLEINPCLENHGGCDRNAECTOTGPNQAVCNCLPKRY--T 411
 465 NKDGTFECLCKDGYEDGDPSSSECRDVNEC-KNPDACGPNQSGCTNTQGGYE-CECLAGFERI 522
 412 GDGKVSLLNVCULTNNGGSPFAFCNVTBODORICTCKRDYTGDDGIVCGSIYGLPKRP 471
 523 AEGAHCTDRDECAVE--PCHPAALCSNTRGSYK-CECRGEGFVGDXTCHH----- 569
 472 STSQYFQLOEHAVRELAQPGFTVPAPLSSSFNHEPRIKMDQGLMSQVLRHYHVGCO 531
 570 -----TILYPIISNDSTVYIPR--SMD----- 587

```

OY 532 QLLDNLKKYTTASTTLOGEFVSVISVSDYVFINNNEAVLSDSDISTNCVHIVDKLSPK 591
Db 588 -----SSTAA---PI-----LSEITF---GKYSKIVLSTII----- 616
OY 592 NLLITPKDALGRVLOMLTTVAANHGYKFKFSKLIDSGLLSVITDSIHTPTVFWPFDKAL 651
Db 617 -----SFDGDLQGL-----IDHAETLKKRAIFAL 640
OY 652 EALPREQODLEFNQDNKKDKLSYLFHVINDSKALASDLPRASAWKTLQSGSELSVRCGTG 711
Db 641 HA---QFDYI-----REGLVAYT---FINDTDAVTLPL-----LMRSSIGIQTAGM 680
OY 712 SDIGELFLEDMOKRFTIHRGLLPDVGVAYGIDCLMLNPITLGGRCDFTFEPDIPGEGCS--- 768
Db 681 I---ENFQTKRLHIF-----TFDRVROSGSENF 705
OY 769 -----CIPTPKCPILSKRKPGYKKCIYNLPERRNVEGCONLCTVYO 811
Db 706 NSFQIVLASSDKDATILSLITLYEKVQARGPMTGISPSRFLMPPNNRLASGSN-----VGQ 760
OY 812 TTRCGHGTFMPDQACPGGDTPCNNRGKRCDLTTPMGQCLCHTGFNGTACELCHMGARF 871
Db 761 PKGMHTRVADIADLOTCTPPG-----RLGEPLCD-----RECAAGHYG 796
OY 872 PDCQPR-SCSEHGQCEGILTSGEC---LCETGWTAAASCDPTFAVFAVCTPACSVNATCT 927
Db 797 INCESTCHGDGSAVCV-ITGM--CPGALCRAGWEGSSCD-----Q 834
OY 928 ENNTCYCNLTREBDGIITCTIVYDFCKONNGCCAKYAKCSQKCTQVSCSKKGYKGDGYSCI 987
Db 835 DIDECEMSTL-----YTCAAGSQCQVNTRG-----YRCDCGFAPVGYKECK 875
OY 988 EIDPCADGVNGGCHEHATCRMTGPKHKCKCKSHYVGDGYDC 1029
Db 876 PIDRCLSRFSVPCSRNAECYESIESNPKCYCKRGYHGDGFC 917

RESULT 9
US-60-360-039-5866
; Sequence 5866, Application US/60360039
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Chen, Xianfeng
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)A
; CURRENT APPLICATION NUMBER: US/60/360, 039
; CURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 5866
; LENGTH: 3051
; TYPE: prt
; ORGANISM: Caenorhabditis elegans
US-60-360-039-5866

```

```

Query Match          4.7%: Score 311; DB 7; Length 3051;
Best Local Similarity 20.4%: Pred. No. 1.2e-12;
Matches 244; Conservative 109; Mismatches 127; Indels 416; Gaps 68;

QY 133 OKNRCDNNT-----IIVGEC---GKSGQAPCPLETKPLRTRCICYSITPM 178
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 818 ENEKCTNDSDFCECLPQFERIRMGOCAPGSCNPNDSMD---VRRKOCL----- 868
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 179 GKRSYFIGQ-----PQCVRT--IITRACWLASTAH 207
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 869 -PRGNIYITCGCGRNEKRPHTDICKNECLTGEHDCDRSARCIDPDESXYICAGSGFIDH 927
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 NAKPAPGEKMKALSTASVMDGVNLTG-----TCQCGIG-----FNCT 245
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

D	b	928	SPNSENPRGRVCAVLAQNECLDGSNRCSPNALCTIDTEEGYVCRKSKSFVDYSPNPOTFFGM	987
Q	y	246	ACE-----TCTGKAGIHCQDQACSVHGRCSQGLFADGSCDCDVGW-----RGVCKDM	293
D	b	988	VCKELVNCTNTPRLIM-QCDRAHAI-----DTIEGYSCIKRGFVDMDFGNPCRBEO	1040
Q	y	294	EITTDNCN---GTCHTSANCLLDEBQKASCKAAGRGN-----GVYC-----T	334
D	b	1041	IKTNKCSPGKNDKDRNAKCIQIIDDYSCACPPGFKDKSPSSRRGRCLIPVIECDNP	1100
Q	y	335	AINCETISNGGCSIKKACKRTPTGNNRVCYCKAKY-----TGOSIVLEI-NPCLN	385
D	b	1101	TLNDDSDP-----RAYCTDIDG-YMCRGQFLDISISVYKPRCLKPLQNECALGI	1154
Q	y	386	GGCBNAECTQTPGNOAVNCMLPKY-----TGQKYS-LINCLTLNNGCSPPAFAC	436
D	b	1155	DDCADDGICEDNDPSTCTCKCANMYLDVSDRYTRGRCKRLINCYQYQNDGSELEATC	1214
Q	y	437	NTTEDQRICTC-----KPYTG-DGIVCRGSIYGLLPKNBSTSYFFLOEHAUREL	488
D	b	1215	TDTE-DSYICACPOSHIDISPTVNRGRKLMRI-----NECSIN-----RHD	1251
Q	y	489	AGPGFFVFAFLSSFFHMERIKD-----WDQGLMSOYLRYHVYCGQQLLDMLKATT	542
D	b	1258	CSPNADCIIDPES-----YKCCCRDFEVDESDDSSRRRPR-----CRPALVDCRCRG	1306
Q	y	543	S-----ATTLOEPAVISVSODTVEINNEAKVLSDDIISTNGVLIHVIDLKSLKNLLIPK	598
D	b	1307	HDCHVNAICQDLPQGYTCQCSADPVD-----VSHRASHP-GRICQPRTPRPPE	1355
Q	y	599	---DALGRVIONLTTVAANH-----YKFSKLIQDSGLLSVITDSTHTPTVFW	645
D	b	1356	CRLDGNGCKHNLNEVCRLMGEPKSCPVNYORDS-----SOSCIINECLFT-----	1404
Q	y	646	PTDKALELTPREODFLFNQDNKDKLSYLKFNIVIDSALASDLPRSAKMTIDGSELS	705
D	b	1405	-----QLNDCHTAAIDCIDIQVGY-----T	1423
Q	y	706	VWCGTG-SDIGEL-----PLNEOMCRFIH-----RGLTFDVGAYVIGDCL-LMN	747
D	b	1424	CQCRGEGFIDIDRRRPRGRMKPWAYNE--COYPLINDCHQNAACIDLEBGTBECKMGMMD	1481
Q	y	748	PTLG--GR-CDTFTTFDIPBEGSCJLFTPKPLSKPKGVKKC-----IYNPLD-----	794
D	b	1482	HSHGPRGRICKQLINECLRLSLNSCDRNARC-IDKEBEGEBCRGFIDVSPPTLKGK	1539
Q	y	795	-FRNVECCQNL-----CTVVIQTPRCCHGFMPDQACP-----GGDTP-----	834
D	b	1540	ACRELNVNCAASRLNDCD--KNAKCKDITMDSYECD-CPVNSKDISPSPSPGRVCLMET	1595
Q	y	835	-----CNRNRCMD--LYTPMOQCLCHNGF-----NGTAC-----ELCMHG	868
D	b	1596	NECSGVHDCDPSATYCKRDNOSF-----CECPGFFVDRSPNKHARPGVCAVLDECREG	1651
Q	y	869	RFGPCDQSRCSSEHQGC--DEGTTGSGBELCTETGTAAS-----CDTPRAYAV	915
D	b	1652	R-----HTGSSHADCRDLEBGT--CECDGTVDSPNLASDPRGVGCAPE-----EV	1696
Q	y	916	CTP-ACSVHATCTE---NNTCVCLNLEBGGDITCTVVPFCQNN-----GGCAKVA	962
D	b	1697	CPPNHDCSSAAVCEPPLGMYKQCVCIQGYDQSPGQKGRVCYRNNAKCHDPLRLNCSNA	1756
Q	y	963	KCSQKGTQVOSVCKKGV-----KGGYSC-----IETDPACDYVNGGCHAHATR	1007
D	b	1757	ICYDEPRGYCECKRGMDSPDSSQGRVCEPRPPSPRRHPCDPERRNDCHPAGTCR	1816
Q	y	1008	MTGPBKHKCECKSHYVGDGVYCEFEQILPLDRCL-----ODNGCHPADSACDL	1055
D	b	1817	ATGAGSYCECLSLSYADRSPD--PRNKGRLCYLTFPVCYCLDPBQNDCHAAALCSEV	1870
RESULT		10		
US-60				039-5971


```
Db 1457 EVSGIYGV-----LLDGLPPLHAHCAFMKSSDVINYGPILSYALEDKDN 1505
OY 560 -----TVPINNAKAYLSDIISNGYIH-----VIDKLSPKNLL 594
Db 1506 TELLIDYNGWLYVNGEKEKTIKPSVN-DGIMHIAITWISIGAMRYIDGELSDGCTG 1564
OY 595 ITPKDAL---GRVLONLTVAANHGYTKFSKLIODSGLSVITDSIHPTVTFMPTKAL 651
Db 1565 LSIQKALPGGALVLDGEODKKEGFPNPAESFVGSISQLN-LMDYVLSPPQV-----KLL 1618
OY 652 EALPPEQ-----ODELFNODNKDKLYLKFHYIRDSKALASDLP--RSASWKTQ 700
Db 1619 ASSCPPELISRGVNLAMPDFLSGITGKVKVSSSMF--CSDPPSLEGSVPHLRPASGNKP 1676
OY 701 GSELVRCGTSND-IG---ELFLNEQM-----CRFIHRGLL-----FDVGV 737
Db 1677 GSKVSLFCDFQFQWVGNPVQYCLNQGOWTOPLPKRCERIGCLPALENGFTSAEDFHAQS 1736
OY 738 AYGIDCLLMPPTLGR---CDTFTFDIPBEGSGCIFTPKCPKLS-----KPKGVK 785
Db 1737 TVTYQCTSGYLLDSDSRMFCXDNCSMN--GISPCLDVBDECAVGSDESHASCLNTNG-S 1793
OY 786 KCLLYNPLPRRNVEGCONLCVYIOTPRCHGYFMPDQACPGGPDTPCANNRGMCRDLY 845
Db 1794 YVCSGNP-PYTGDKKNCAE-----FVKCK-----APENDENGSR-SGEIV 1831
OY 846 TPMGQCL---CHTGFN--GTACELC-----WHGRFPDCCOPRSCS-----EHG----- 883
Db 1832 T-VGTAVTFSCDEGHVLYGVSTITCLETEGW-DRLRPSCEAISCGVPPVPGVGDGSAF 1889
OY 884 -----QCDEGITGSGE---CLCETGWTASCDPTTAVFAVCTPACSVAHATCTENNT 931
Db 1890 TYGSKVYVRCGKGYTLSDDEESACLASGWSHSS-----PYCGI-VKCSQED 1936
OY 932 CYCNLANTEGOGITCTVDFCQNNNGGCAKAKVCSQKGTQVSCSKKKGKGGYSCIEIDP 991
Db 1937 -INNGKTYLSGLYVLSI-----ASYSCENGYSLOGPSLLE--- 1970
OY 992 CADVNGSGCHEHA--TCRMTGPGKHKC-----ECKSHYVGADV-- 1028
Db 1971 C---TAGSWMRARPSCOLVSCGEPYIKDAVITGNSFTFPGNTVAATYCKEYITLAPRTI 2027
OY 1029 -CEPEOLPLDRCLDODNGCNPDASCADLYFODTIVGVFHLRSPLOYKLYFDKAKACAK 1087
Db 2028 ICQAN---GKWNSSNHQCLA-VSC-----DEPPNVDH-ASPETHRLFDFTAFFUYCA- 2074
OY 1088 EMATITATYNOLSTYOKAKYHICSA--GWLBSGRAVYPTTYA--SOKGCAVNVGIYDYSR 1143
Db 2075 DGYSGLADNSOL-----ICNMQGNVPPAGQAVPRCIAHFCCKPPSVSYSLIESYK 2125
OY 1144 ANKSEMDVFCYRMKDVNCTKAGYV-----GDGFSGS-----GNLQ 1181
Db 2126 A-KRAAGSVYSFK-----CMEGFVLNTSAKIECLRGEMSPSLVQCIPIVRCGEPPS 2177
OY 1182 VLMSFPLTNE-LNEVLAFS-----KSS--ARGQ----- 1207
Db 2178 IANGYPSGTNSFGAVVAYVSHKGFYIKGEKSTCEATGQSKSPRTCHPYSCNBPVKE 2237
OY 1208 -AFLKHLTDLISR-----GTLFV-----PONGSLP---G 1232
Db 2238 NGFLEHTTGRTFSEARFOCNPGYKAAGSPVFCOANRHMHSADPLSCTPLNCGKPPETQ 2297
OY 1233 NKSLSGRDIE-----HHLTNNVSGFYND-----LVN 1258
Db 2298 NGFLKGESEFEGSVQVQFCVNGYELVGNDSWTCQSGKWSKPSKPCVPTKCAEPPLLEN 2357
OY 1259 GTFPLRTMLGSOLLTFPSODOLHOETRFVDGRSIL-----OMD----- 1295
Db 2358 QLVKELASLEVWNTISKKEGHA---LQGPSVLKCLPBGQWNGSFPLCKMVLCPSPPLI 2413
OY 1296 ---IIANGLIHISEP-----LRAPPTAATAHSGLGTEGI-FCAYV-----LV 1335
```

```
Db 2414 PRGVASSGALHFGSTYKYLICVDGFELKSGPTLICOADSTWSSPLPECVPEVCEQPEIL 2473
OY 1336 TGAIALAASYFRLLKORT--TGFORFDQKRTLM-----SMLASSSPRISQTLKMRPQ 1386
Db 2474 NGIIHVGLATLSTTLYTCRKRGFELVGNATILCGENQWL--GSKPMKPIECEBP 2528

RESULT 14
PCT-US02-09671-698
; Sequence 698, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: Zycos Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026WO1
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 698
; LENGTH: 3623
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-698

Query Match 4.3%; Score 338; DB 1; Length 3623;
Best Local Similarity 31.4%; Pred. No. 9.9e-11;
Matches 101; Conservative 30; Mismatches 125; Indels 66; Gaps 21;

OY 182 SVFIQCQPCQRCVTRITTRACWLASLAHNKRAPGEVKMKGALGASVWDVNCNCTGCGGIG 241
Db 153 SFCICPPQ-----WKGPL-----CSADVNECEIYSGTFLSCQNG-GTCVNTMG 195
OY 242 FNGTACERCTEGKXYIHC-----DQACSVHGRC-----SGGPLGDSQCDQVGM- 286
Db 196 ---SVSCH-CPEYITGPQASKYDCEGGSVAVRCHVIGICEDLMREBAGRPKXSCVCDACMM 252
OY 287 ---RGVKDMETITDNCN---GTCHTSANCLLPDQKASC-KCAAGFRNGTVCATAIAC 339
Db 253 SSPNSPACTLD--RDECSFQPGPCSTLVQC-FNTQGSFYCGACPTGMOGNGVICEDINEC 309
OY 340 ETSNGCGST--KADCKRTTPGNRV--VCKAGYTGIGIYCLEINPLLENHGGCDRRAEQTQ 396
Db 310 EINNCGCVAPPVREV--NTPSSSHQACPPYQGGRGVCTLTLDICSVSNGGCHPDASCS 368
OY 397 TGNPNQAVNCLPKYTGDS---KYCSLINVCLTN---MGCSFPAFCVYTBODORICTCK 449
Db 369 TLGSLPCTCLPGYTGNGYGPNGCYQLSNICLSHPLCLNGQCID-TVSGY-----FKACD 421
OY 450 PDYTGDIYCRGSIYGLPKNP 471
Db 422 SGWT--GVNCTENI-NECLSNP 440

RESULT 15
PCT-US02-09671-701
; Sequence 701, Application PC/TUS0209671
; GENERAL INFORMATION:
; APPLICANT: Zycos Inc.
; TITLE OF INVENTION: TRANSLATIONAL PROFILING
; FILE REFERENCE: 08191-026WO1
```



```
; CURRENT APPLICATION NUMBER: PCT/US02/09671
; CURRENT FILING DATE: 2002-03-28
; PRIOR APPLICATION NUMBER: 60/279,495
; PRIOR FILING DATE: 2001-03-28
; PRIOR APPLICATION NUMBER: 60/292,544
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: 60/310,801
; PRIOR FILING DATE: 2001-08-08
; PRIOR APPLICATION NUMBER: 60/326,370
; PRIOR FILING DATE: 2001-10-01
; PRIOR APPLICATION NUMBER: 60/336,780
; PRIOR FILING DATE: 2001-12-04
; PRIOR APPLICATION NUMBER: 60/358,985
; PRIOR FILING DATE: 2002-02-20
; NUMBER OF SEQ ID NOS: 2041
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 701
; LENGTH: 3623
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-09671-701
```

```
Query Match          4.3%; Score 338; DB 1; Length 3623;
Best Local Similarity 31.4%; Pred. No. 9.9e-11;
Matches 101; Conservative 30; Mismatches 125; Indels 66; Gaps 21;
```

```
OY 182 SVEICGPOCVRTITTRACWLASLHNAKPARGEVKMCALGTASVWDGVNGTGCQCGLG 241
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 153 SFECICPPQ-----WKGPL-----CSADVNECEIYSGTPI,SCQNG-GTCVNTWG 195
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 242 FNGTACETCTEGKYIHC-----DOACSCVHGRC-----SOGPLDGSQDCDVGW- 286
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 196 --SYSCH-CPPEYTPQCAASKYDCEGGSVARCVHICEDLMREQAGEPKYSCVCDAGWM 252
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 287 ---RCVKCDMETTDNCN---GTCHTSANCLLDPPDGKASC-KCAAGFRNGTVCCTAINAC 339
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 253 SSPNSPACTLD--RDECSFQPGPCSTLVQC-FNTQGSFYCGACPTGMOGNGYICEDINEC 309
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 340 ETSNNGCGST--KADCKRTTPGKRVG-VCKAGYTGDIIVCLEINPCLNHHGCDRNAECTQ 396
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 310 EINNNGGCVAPPVCEV-NTPGSSHQACPPGYGDDGRVCTLTIDICSVSNMGCHPDASCSS 368
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 397 TGPNOAVCNCLPKYTGDG---KYCSLINVCLTN--NGGCSPAFCNVTEDORICTCK 449
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 369 TLGSLPLCTCLPGYTGNGYGPWGCVLNLSICLSHPCLNQCID-IVSGY-----FCKCD 421
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 450 PDYTGDIIVCRGSIYGEI,PKNP 471
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 422 SGWT--GVNCTENI-NECLSNP 440
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: June 17, 2002, 12:34:07
Job time: 313 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 17, 2002, 12:25:19 ; Search time 3351.98 Seconds
(without alignments)
18949.000 Million cell updates/sec

Title: US-09-842-930A-1
Perfect score: 4706
Sequence: 1 tcttaccagctactcac.....aagcaaaaaaaaaaaaaa 4706

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 674847542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:
1: em_estdb:*
2: em_esthum:*
3: em_estnu:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vrt:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	524.6	11.1	664	9	BB626584
2	447.8	9.5	462	10	BB626584
3	377.2	8.0	500	9	AA821602
4	357.6	7.6	758	10	B1838464
5	304.6	6.5	494	10	BF601840
6	295.8	6.3	385	9	BB842118
7	287.6	6.1	459	10	BB656181
8	253.2	5.4	459	10	BB656181
9	248.6	5.3	670	9	BB086042
10	245.4	5.2	1066	10	B1146461
11	243.8	5.2	494	9	BB735524
12	242.2	5.1	439	9	BB847553
13	213.8	4.5	793	10	B1832486
14	213.2	4.5	474	9	BB745695
15	204.4	4.3	373	10	BB745695
16	200	4.2	443	9	AA275305
17	198.4	4.2	460	10	BB447169

18	192.8	4.1	621	10	B1518863
19	190	4.0	809	9	AU080852
20	184.4	3.9	520	10	BG712246
21	179	3.8	378	9	AT852814
22	175.6	3.7	325	9	AA257161
23	175.6	3.5	928	10	B1906024
24	156.6	3.3	516	10	BE808086
25	153.2	3.3	245	10	B1183898
26	153	3.3	317	10	BF660624
27	148.8	3.2	570	10	BG383208
28	147.8	3.1	880	10	B1655595
29	141.4	3.0	405	9	AA138267
30	139	3.0	539	10	BG803514
31	136.4	2.9	460	10	R97483
32	127	2.7	425	9	AT272209
33	125.6	2.7	282	10	B1133987
34	123.2	2.6	634	10	BG964207
35	122.8	2.6	401	10	B1006218
36	122.8	2.6	441	9	AA389617
37	122.4	2.6	527	12	AZ030878
38	122.2	2.6	490	12	AZ487818
39	117.8	2.5	397	9	AT606926
40	117.4	2.5	647	9	AT598662
41	114.2	2.4	619	10	BG712275
42	113	2.4	570	9	BE233324
43	108.8	2.3	371	10	BM148089
44	108.8	2.3	380	12	AZ487512
45	108.8	2.3	475	10	BE479875

ALIGNMENTS

RESULT 1
BB626584
LOCUS BB626584 RIKEN full-length enriched, adult male drosophila mus
DEFINITION Musculus CDNA clone 9330210123 5', mRNA sequence.
ACCESSION BB626584
VERSION BB626584.1 GI:16464521
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 664)
AUTHORS Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hiramoto,K., Horii,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M., Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toyota,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshinori Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayashizaki,Y., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagui,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanabe,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.

library is a subtracted library derived from the following tissues: thalamus, cerebellum, hypothalamus, medulla, pons, midbrain, cerebral cortex, corpus striatum, testis, and hippocampus. For a detailed description of the library from which this clone was derived, please visit our web site at ratelc.eng.uiowa.edu. The subtraction has been previously described in (Bonaldi, Lennon and Soares, Genome Research 6:791-806, 1996)

TAG_LIB=UI-R-CAO
TAG_TISSUE=medulla
TAG_SEQ=GAACCG

BASE COUNT 115 a 104 c 125 g 118 t
ORIGIN

Query Match 9.5%; Score 447.8; DB 10; Length 462;
Best Local Similarity 99.6%; Pred. No. 3.9e-88;
Matches 449; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 4203 ggagacagatctgagacagacacccctcgggggacactcgtctgacatgagaag 4262
|||||
Db 462 GGAGACAGGATCTGGAGACAGCACCCCTCGGGGGCAGTCGGCTCGACATGAGAAG 403
|||||
QY 4263 ccaagcaagcaaccacagtcacggtccacggtatcccaagcccaagctgtctcattggaat 4322
|||||
Db 402 CCAGCAAGCAACACAGTCACGCGTTCCAGCGTATCCAGGCCCGCAGCTGCTCATGAT 343
|||||
QY 4323 cagtggtttaagaatgacacactcataagaacagacatacctcattcgtgtaaat 4382
|||||
Db 342 CAGTGTGTTTAAAGAAATACAAACACTTAAGCCAGCCATACCTCACCCTTGCTGTTAAT 283
|||||
QY 4383 ctggagctgcgcagagagtaagagacatgtctcgtgatacctcggggagacctccact 4442
|||||
Db 282 CTGGATGTTGCGCCCGGCGTAAGGAGCCATGTGCTGGATACCTGGGGAGACCTCCACT 223
|||||
QY 4443 cctctgaagcctataccgtggtctctcctcactccataatggtgctgtctgtccctc 4502
|||||
Db 222 CCTGTGACCTATACCGGTGCTCTCACTTCATATGCTGCTGCTGCTGCTGCTGCTG 163
|||||
QY 4503 tctgtaccacaaactgtactcgtgtgattcctcattgaagtaagacccaagcg 4562
|||||
Db 162 TCTGTATACCCCAAACTGTGACTGTGTGTAATTCCTATTACGTAAACCAAAAGGGG 103
|||||
QY 4563 gggctcaccctctatgtctgtatccagttaccagaagtaaccgccaacatctgac 4622
|||||
Db 102 GGGCTTACCTTTATGTTCTGATTTCAGTAACCAAGAACTGCTGCCACACATGTTGC 43
|||||
QY 4623 tcaataaatgtttggaacaataaaga 4653
|||||
Db 42 TCAATTAATGTTTGGGAACAATTAATAAAA 12
|||||

RESULT 3
AA821602 500 bp mRNA linear EST 17-FEB-1998
LOCUS vva0c03.r1 Soares.mammary.gland_NbMNG Mus musculus cDNA clone
DEFINITION IMAGE:1245316 5', mRNA sequence.
ACCESSION AA821602
VERSION AA821602.1 GI:2891470
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 500)
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mousedest@wustl.edu
This clone is available royalty-free through LIND; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:659004
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 465.

FEATURES
source
1..500
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1245316"
/clone_1ib="Soares.mammary.gland_NbMNG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pT73D-Pac (Pharmacia
) with a modified polylinker; Site_1: Not I; Site_2: Eco
RI; 1st strand cDNA was primed with a Not I - oligo(dT)
primer [5',
TGTATCCATCTGAAGTGGGAGCGCGCGCAATGGTTTTTTTTTTTTTTTTT
T 3']; double-stranded cDNA was ligated to Eco RI
adaptors (Pharmacia), digested with Not I and cloned into
the Not I and Eco RI sites of the modified pT73 vector.
RNA provided by Dr. Minoru Ko, Wayne State Univ. Library
constructed and normalized by Bento Soares and M.Fatima
Bonaldi."

BASE COUNT 129 a 141 c 112 g 118 t
ORIGIN

Query Match 8.0%; Score 377.2; DB 9; Length 500;
Best Local Similarity 87.9%; Pred. No. 1.4e-72;
Matches 434; Conservative 0; Mismatches 56; Indels 2; Gaps 2;

QY 1275 gaacatagcgagctgcagtcacatttgctctcgaactacactgagcaagatcaagat 1334
|||||
Db 9 GAATTAAGGGGGCTG-AGTCCATTGGCTTTCGATACACAGAAACAAGATCAAAAGC 67
|||||
QY 1335 atgtacctgcaagcagactacacgggtgataggaaatgctctgacggggcagcatcagg 1394
|||||
Db 68 ATGTACTGTCAAGCCAGACTACACGGGTGATGGAATCGTCCCGTCGACGATCATTC 127
|||||
QY 1395 ggaagctcccaagaacctgagcgtcccaagtaactcttcagagttgcaagagatgctg 1454
|||||
Db 128 GGAACCTCCAAAGAAATCCCTCAACGTCAGATATTTCTTCCAGTTGCAAGACATGCTCT 187
|||||
QY 1455 ccgaagctctgacacccctgaccccttcacggtgttcgaccccttgctcagctccctcaa 1514
|||||
Db 188 CCAAGAGCTTGGTGGACCTGGGCCCTTCACCGGTTGTCCTTCATCTGACCTCTTCAA 247
|||||
QY 1515 tcatgagcccgagatlaaagactggatcagcagggcctcattgtcccaagttcttcgcta 1574
|||||
Db 248 CAGTGAGTCCAAGCTTAAAGCTGGGATTAACAGGGCCCTCATGTCTCCAGATTCTAGGTA 307
|||||
QY 1575 tcaagtggtgggtctcccaagcagctgctgttggaacaacctcaaaagtagcaagaagtgccac 1634
|||||
Db 308 TCACTGGTGGCCTG-CACACACTACTGCTGGAGAAACTTAAGGATATCAGACAGTCCAC 366
|||||
QY 1635 gaccctccaagaagagccagttccatctctgtctctcaggaacactggtgtcataaaca 1694
|||||
Db 367 AACCTGCCAAGAGAGAGCAATTTCCATCTCTGTCTCTGAGACACTGCTCTATAACAA 426
|||||
QY 1695 tgaagcgaagctctcgtccagtgacatcatcagcaacaaatggcgtcgtccatcagttataga 1754
|||||
Db 427 GAAGGGAAGGTTCTGTCCAGTACATTAATACAGACCAACGAGATCATCAGCTCATAGA 486
|||||

Y	175	caagttgctgtctc	1768
Db	487	CACATTGTTGTCTC	500
RESULT	4		
BI838464			
LOCUS	BI838464	758 bp	mRNA
DEFINITION	603083445F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5222306 5',		linear EST 04-OCT-2001
ACCESSION	BI838464		
VERSION	BI838464.1		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 758)		
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC)		
	Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: cgapbs.rem@nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	clone distribution: MGC clone distribution information can be		
	found through the I.M.A.G.E. Consortium/LLNL at:		
	http://image.llnl.gov		
	Plate: LRAM11559 row: 3 column: 11		
FEATURES	High quality sequence stop: 737.		
Source	Location/Qualifiers		
	1..758		
	/organism="Homo sapiens"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5222506"		
	/clone_lib="NIH_MGC_120"		
	/lab_host="DH10B"		
	/note="Organ: pooled pancreas and spleen. Vector:		
	pcwv-SpOrF6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA		
	source anonymous pool of spleen and pancreas from 28 yo		
	male. Library is oligo-dT primed and directionally cloned		
	(EcoRV site is destroyed upon cloning). Average insert		
	size 1.5 kb, insert size range 1-2.5 kb. Library is		
	normalized and enriched for full-length clones and was		
	constructed by C. Gruber (Invitrogen). Research Genetics		
	tracking code 025. Note: this is a NIH_MGC Library."		
BASE COUNT	195 a 182 c 224 g 157 t		
ORIGIN			
Query Match	7.6%; Score 357.6; DB 10; Length 758;		
Best Local Similarity	77.1%; Pred. No. 3.1e-68;		
Matches	563; Conservative 0; Mismatches 154; Indels 13; Gaps 10;		
Y	446	gtggaagtggtccaggaagcccccgtccacatcgagacaaccact--tagagaga	502
Db	1	GTAGGACATGCGCTCCAGAGCTGACCTGCCCTTCGGAACCTAAATCTCTAGGTAATGAGA	60
Y	503	cgaggaatgcatctatccatctactctcatgtygggaagagatccglatatcatcggtgac	562
Db	61	AGAGGAGATGCATCTATACCTCTCTTTTCATGGAAGACGAACCCCTGTTATTGGGTGCC	120
Y	563	agccacagtggtggaagaccatcatcaagaagagctgctg-9ctggtctcttggtccac	621
Db	121	AGC--CAATGTGTGGAACCGTCATTTAGCAGAGAATGCTGTCGCCGCTTCTTTGGCCCC	178
Y	622	aatgcgaagctgcgcccggaagagtlcaaaaagtgtgctctgtygaagaggtctctgtcgtg	681
Db	179	AATGCCAGCCCTG-TCCAGGAATGCCAGAAATGTCTGCTTGTGTAATGSCAATCTGT-TTG	236
Y	682	gaacgtgtgtaatgtagctgcaagtgcagctgacgtgaggtctgaggtctcaatgtaggaagcgtgt	741

Db	237	GATGAGTGAATGGCAACAGCTGTGTGTGAAGTGTGGGAGAGGGCTTCAGGGCAACAGCCTGC	296
QY	742	gaaacctgcacctgaggggaagatlglatlaccactgcgcgaaccaagatgctctgtgtcat	801
Db	297	GAGACCTGCACCCAGGGCAGTACGGCATCCATCTGTACCCAGCAAGCATGTTCTTGTGTCCAT	356
QY	802	gggagatgtagcgaagacccttgggaagagtgctctgtgactgtgtagctggagctgggca	861
Db	357	GGGGAATGCACCAAGGACCCCTTGGGAGATGCTCTGTGTACTGTGATGT-6GCTGGGCA	415
QY	862	ggaatgaagtgtgacatcgagatcacacagaacaactgaacggaacgcgtcacaccagt	921
Db	416	GGATGCGATTGTGCATATGCACACACAGAACCACTGCATATGGAGCATGGCCATTCCACGC	475
QY	922	gccaaactgtcctctgtagtcacagaagcgaaagcctgtgccaatgtgcgcaggaattccga	981
Db	476	GCCAACTGCTCCACCAACTCAGATGGTAGACGCTTCAATGCAAGTGTCCACAGGATTCAC	535
QY	982	ggggaatggaacggtctgcga-caagcatcaatgtctgtlgagacag-caatggagatggt	1033
Db	536	GGAAACGGGACCATCTGTCCACCAACATCAATGCGCTGTGAGATAGGCCAATGGAGGTGTGT	595
QY	1040	ctcaaaagggccgaactgtcaaaagacaaccaccaggaacccgagtggtgtgtgcgaagca-	1098
Db	596	CTGCCAAGGCTGACTGTATAGAGAACCA-CCGAGGAAGGGGAGGTGTCCAGTGTGCAAAAGCAG	654
QY	1099	ggctataccggcgagacggtacgtgtgtgcttgaatacgaaccggtgttlttgagaagcaatggt	1155
Db	655	GGCTAACACGGGTGATGTGCATTTGTGCTTGGAATACACCCGTGTTGGAGACATGTGT	714
QY	1159	ggcgtgtgaca 1168	
Db	715	GGCTGTGACA 724	

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	MEDLINE	COMMENT
BF601840	2669924 MARC 3BOV Bos taurus cDNA 5', mRNA sequence.	BF601840	BF601840	BF601840.1	GI:11699063	Bos taurus	EST.	Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahnenrugg,S.C., Bennett,G.L., Heaton,M.P., Laegreid,M.W., Rohrer,G.A., Chitko-McKown,C.G., Petee,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and Keele,J.W.	Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle genome Res. 11 (4), 626-630 (2001)	USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smitht@mail.marc.usda.gov	21180013	Contact: Smith TPL USDA, ARS, US Meat Animal Research Center PO Box 166, Clay Center, NE 68933-0166, USA Tel: 402 762 4366 Fax: 402 762 4390 Email: smitht@mail.marc.usda.gov
FEATURES	Seq primer: ATTTAGGTGACACTATATAG.	PCR primers	FORWARD: AGGAACAGCTATGACCAT	BACKWARD: GTTTCACGATCAGCAG	Plate: 42	row: C	column: 11	and -mismatch 12 options.	Single pass sequencing. Bases called and alt. trimmed with phred v0.980904.e. Vector identified by cross_match with the -mismatch 18	Location/Qualifiers		

```

source
1. 494
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 3BOV"
/rissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
library made from pooled tissue from marrow, alveolar
macrophage, ovary, fetal semitendinosus muscle, and fetal
longissimus muscle."
BASE COUNT      130 a      122 c      132 g      110 t
ORIGIN

Query Match      6.5%; Score 304.6; DB: 10; Length 494;
Best Local Similarity 80.0%; Pred. No. 1.4e-56;
Matches 396; Conservative 0; Mismatches 94; Indels 5; Gaps 3;

QY 205 gtccctggggaagaagctcctgaagaatgactgcatgaacgcatgcaccgagagacatg 264
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTCTTAGAGGAGAAAGCTCTGAAGATATACCTGCACAAATGGCATGCACCGGAGACCATG 60

QY 265 ctgggggtctccctacccctctgctctctctccgcgaatgacacagctglatgtaaatgaa 324
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTGGGCTTCTCTACCTCTCGGCTTTCTTCCCTCATGATGACCACTCTATGTAATGAG 120

QY 325 gctccaaataactacacaaatgtggccactgataaaggatgatacgtgctctggagaa 384
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 GCTCGGATTAACCTACACCAATGTAGCCACTGACAAAGGAGATGATCATGGCTTGGGAAA 180

QY 385 gtctcgaatcagaagaacagatgtagacaataatgacacattatgtgagaggaggag 444
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GTGCTGGAATTCACAGAAATAGATGATGACATTAACGACACTACTACTGATGAGAGAAAG 240

QY 445 tctggaagaatgtcccaagaagccccctgcccactcgaagacaaacacactag---aag 501
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TGTGGAAGTGTCCCGACCCCATCTGCCATTTGGAACTAAGCCATGAGCCACAGAG 300

QY 502 acggaggaatgcatctatctcatctcaactcattcgtgggaagagatccgtattcaatgg 561
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 301 ATGAGGAGATGCTCTATATACCATTTATTTCATGGGACGACGATCCCGCTTCTTAAGGTGC 360

QY 562 caagcagatgtgtgagaacatcattacaagaagcctgtg-gctgctctctctggccca 620
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 361 CAGCCAAAGTGTGTAAANAACGCTCATTTACAGAGAAATGCTGCGCGGCTACTTTTGCCCC 420

QY 621 caatgccaaagcctgccccgggagaggtcaaaaatgtgtcctctgggaagcgtctgtctg 680
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 421 CAGTGCACGCTTGCCCGCAGGAAACCTGAGAAACGTCTGCTTGGAAATGGGATCTG-CTT 479

QY 681 ggaaggtgtgaatgg 695
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 480 CGACGAGCTCAATGC 494

RESULT 6
BB842118      385 bp      mRNA      linear      EST 21-NOV-2001
LOCUS      BB842118      RIKEN full-length enriched, 6 days neonate spleen Mus
DEFINITION      musculus cDNA clone F430003K15 5', mRNA sequence.
ACCESSION      BB842118
VERSION      BB842118.1 GI:17042849
KEYWORDS      EST.
SOURCE      house mouse.
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 385)
AUTHORS      Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
Hayatsu,N., Hiramoto,K., Hiraoaka,T., Hirozane,T., Imcitant,K., Ishii
Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Saito,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,

```

```

TITLE
JOURNAL
COMMENT
Shibata,K., Shingawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
,A., Takahashi,F., Takeku-Akashira,S., Tanaka,T., Tomaru,A., Toya,T.,
,Watshiki,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshinide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp,
URL:http://genome-gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
wag1,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multipillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome-gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source
1. 385
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="F430003K15"
/clone_lib="RIKEN full-length enriched, 6 days neonate
spleen"
/rissue_type="spleen"
/dev_stage="6 days neonate"
BASE COUNT      88 a      97 c      110 g      90 t
ORIGIN

Query Match      6.3%; Score 295.8; DB: 9; Length 385;
Best Local Similarity 87.3%; Pred. No. 1.2e-54;
Matches 336; Conservative 0; Mismatches 47; Indels 2; Gaps 1;

QY 2453 gtacttcacgaacactgtcagagctgccccctggagagacagatacacacgtgttaaacac 2512
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1 GTTACTTATGACGACACTGTCTAGCGCTCCCTCGAGGACACAGACACACCTGTATACAAAC 60

QY 2513 ggggcatgtgcccgagatctgtacacaccatgaggacagtgccctatgacacacggcttca 2572
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 GGGCATGTGTTATGATCATGACAAACCACAGCAGACAGTGCACCAATGATCATCTGTTTCA 120

QY 2573 acggagacagcctgagatctgtgtgcatgtggagatgtggcgctgtagctgtgaacccgca 2632
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 ACGGACAGCCCTGTAGCTCTGCTTGGCTGGAGATTGGTCTTGACTGTCAACCTCTTG 180

QY 2633 gctgtccgagcatgagaagtgtagtagaggatcacagagctccggaggagtgccctgtg 2692
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GTTGCTTGAGAGATGGGAGATGTGACGAGGAGATACAGGCTCGGAGCATGSCCTCTGTG 240

QY 2693 aa--acagagtgtagacgcgcgtctgtgtgacactccacagagctgatactcgaagtgtgac 2750
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 AAGAGCAGGGTGGAGACAGCGCGCTTCTGTGATGATCCACAGTTTGATTCAGTGTGCAT 300

```


FEATURES	Location/Qualifiers
Y 4037	actccgcctaaagcagcgaaccactggttcttcacgctttgatacgaagagagacttga 4096
Db 59	ATTGGCGCGCTAAACACAGACAGCAATGGGATCCGGCGTTTAGTCAAAGATTAACATGA 118
OY 4097	tgctctggccttttggcaagcagcagcagcccaagaatactgcaaaacctctgtatgaacctc 4156
Db 119	GGCCATGGCGCTTGACAGACAGCACGCTTAAAGTAT--GCACGGCCGGTATACAGCGTA 176
OY 4157	agcgcgcgcacccccagagctctctctgtgaccccttcacagacaccttgagaacagatct 4216
Db 177	AACGGGGGACAGCACAGGAGCCGCTCTGGGCCCCCTGCACAACTTTGGGAACGAGAGCT 236
OY 4217	gaagagacgcagccctctggggggaacctgcgctccctgacacttgagaagcagcaagaacca 4276
Db 237	GGAGAACAGATGACCTCTTGGGGCACTGAGGTCCTGATGTGAAAGCCAGCCACAGCCA 296
OY 4277	cag-----tcaagcttccacggtga-ttccagaccccaagctgtctc 4316
Db 297	CAGCCACGCTGACCCCTGAGCCATCACGCTTAACGCTGACTCCCTGCTCCACGCTGTAC 356
OY 4317	atgagctcagttgttlttaagaaatgacaactcataaagcagacataccctctcgt 4376
Db 357	TTGGATCATTTTGTTTTAAAGTATGATTAACACTGAG-----AGCCATACCTCACCCCTCG 410
OY 4377	gttaatctggagattgtgcgcagaggtcgaagagcaatgttgccttgataactctgaggagact 4436
Db 411	GATATCTCTGAGATTTTGTCTGCATGGGTAAAGGAGACATGTTTACAGGACACAC--AGTATCCT 469
OY 4437	ccaactcctctgagcctctaacgctgtctctc--cacttcacatgagtgactgtgtct 4494
Db 470	CTGCCTCTCATGGGCGCTTACACTGTGTGCTCTTATGCACCTTCCAAATGGTGCTTGGCCTGTT 529
OY 4495	ctgcctcctctgttaccacaacaactgtgactctgtgtatcttcctatgacgaataagac 4554
Db 530	ATGCCCTATCT----ATGGGTAGCTGTGACCCCTGGTATCTCTATTAACTGTAAAGCAC 585
OY 4555	caaaagcgggggttactactctatgtctctgtatcttcagagaccagaagtactgtccacac 4614
Db 586	CAAAAGCAAGGCTTATCCCATATGTCTGTATTTCCAGTACCCAAAGATCTGCGCACAC 645
OY 4615	atgtgtctcataaatgt 4633
Db 646	GGGTGCGCTCAATTAATGT 664
RESULT 10	
LOCUS	B1146461 1066 bp mRNA linear EST 05-JUL-2001
DEFINITION	60291356071 NCI_CGAP_L19 Mus musculus cDNA clone IMAGE:5054630 5',
ACCESSION	B1146461
VERSION	B1146461
KEYWORDS	EST.
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE	1 (bases 1 to 1066)
JOURNAL	NIH-MGC http://mgc.nci.nih.gov/ .
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strusberg, Ph.D. Email: cga9pbs-remail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: Plate: LLAM1148 row: g column: 15 High quality sequence stop: 492.

Source	1. 1066	/organism="Mus musculus"	/strain="FVB/N"	/db_xref="taxon:10090"	/clone="IMAGE:5054630"	/clone_1ib="NCI CGAP L19"	/lab_host="DH10B (T1 phage-resistant)"	/note="Organ: liver; Vector: pCMV-Sport6; Site: 1; Notif: Site:2: Salt; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.9 Kb. Constructed by Life Technologies. Note: this is a NCI CGAP Library."
BASE COUNT	272 a	330 c	276 g	188 t				
ORIGIN								
Query Match	5.2%	Score 245.4;	DB 10;	Length 1066;				
Best Local Similarity	80.2%;	Pred. No. 1.7e-43;						
Matches	389;	Conservative	0;	Mismatches 81;	Indels 15;	Gaps 8;		
OY	2358	gaagtgatctacaacccgttacccttcaggaggaacgtggaagctgccaga-acctgt	2416					
Db	1	GAGGTGATCTCAACACCCCTTACCTTTCAGGAGGCGATGTGGAAGCGTTCACACACTGT	60					
OY	2417	gaacggtgtgttcaaaccaccagtgctgc--catgttacttcatgcagactgtca	2474					
Db	61	GCACCTGGTGTCCATGTCGCCCGGTGTGACCTGCTTACTTACATGCCAGACTGTCA	120					
OY	2475	ggcctgccttgagagaca-gataacgctgtaacaacccgggcatgtgcgcgactcgt	2533					
Db	121	GGCTGTCCCTGGAGAGACACGACACCGCTGTACACACCGGGGCAATGTATTGATCACT	180					
OY	2534	acacacccaatggagacagtgctatgccaacccggttcaaaggagacagctcgtc	2593					
Db	181	ACAAACCCACAGACAGTCCCAATGATCATACGATTTCACACGGGACAGCCTGTGACTCT	240					
OY	2594	gctgcagatggagatattggacctgactgtcagcccccagcgtgcctccagaga--tgaca	2651					
Db	241	GGTGTGCTGGGAGATTGTGCTCTGACTGTCAACCCCTGTGCTGTCTGTGAGCACTGGCGCA	300					
OY	2652	gtgtgtagagagatcacagagctccgggga-gtgcctctgtgaacaagggtgagacagc	2710					
Db	301	GTGTGACGAGGGGATATCACAGGCTCCGGCGATGTGCTCTGTGAGAGGAGGTGACAGGCC	360					
OY	2711	cttcgtgtgacactccacagctgtt-atcgcagtggtgacacactgtctgtc-----cg	2764					
Db	361	GGTTTGATGATCTCCCAAGTGTGTATTCAGAGTGTGATCATCTGTGCTGCCGTGCA	420					
OY	2765	tgcagcgccacctgtacggagaacaacagctgtgtgtgtac--tgaactaagaagtta	2822					
Db	421	CAGCCACATGCTATGAGCAGAACACATGTTGTGTATCACTTGCAATTTCAGAGAGTGA	480					
OY	2823	caggga 2827						
Db	481	CGGAA 485						
RESULT 11								
LOCUS	BB735524	494 bp	mRNA	linear	EST 15-OCT-2001			
DEFINITION	BB735524 RIKEN full-length cDNA clone F430005K15 3, mRNA sequence.							
ACCESSION	BB735524							
VERSION	BB735524.1	GI:16134674						
KEYWORDS	EST.							
SOURCE	house mouse.							
ORGANISM	Mus musculus							
REFERENCE								
AUTHORS	Akimura,T., Arakawa,T., Carinci,P., Furuno,M., Hanasaki,T., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K., Ishii,Y., Ito,M., Kawai,T., Kojima,Y., Konno,H., Kouda,M., Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Nunasaki,R., Okazaki,Y., Okido,T.							

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
(bases 1 to 474)
Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanaaki,T.,
Hayatsu,N., Hiramoto,K., Hirooka,T., Hirozane,T., Imotani,K., Ishii
, Ito,M., Kawai,J., Kojima,Y., Kono,H., Kouda,M., Matsuyama,T.,
Nakamura,M., Nishi,K., Nomura,K., Numasaki,R., Okazaki,Y., Okido,T.,
Salto,R., Sakai,C., Sakai,K., Sakazume,N., Sasaki,D., Sato,K.,
Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagawa
, A., Takahashi,F., Takaku-Akhiba,S., Tanaka,T., Tomaru,A., Toy,T.,
Watanishi,A., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@sc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
, M., Kono,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
Wagii,K., Fujiwaka,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watabiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura
, S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Kono,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
, Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
source
Location/Qualifiers
1..474
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="F53009M16"
/clone_id="RIKEN full-length enriched, adult male kidney"
/sex="male"
/tissue_type="kidney"
/dev_stage="adult"
/lab_host="SOLR"
/note="Site_1: XhoI; Site_2: SstI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAGCGCGCCGCACTCGAGTTTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence [5'
GAGAGAGAGAGATCCAGAGCTCAATTAAATTAAATTAACCCCCCCCC 3'].
cDNA was cleaved with XhoI and SstI."

BASE COUNT
ORIGIN
101 a 135 c 119 g 119 t

Query Match 4.5%, Score 213.2; DB 9; Length 474;
Best Local Similarity 76.5%; Pred. No. 1.9e-36;
Matches 364; Conservative 0; Mismatches 78; Indels 34; Gaps 7;

4181 ctgtgacccctcacagacccttgagaaacagatctggaggagcagcccttggggc 4240
|||||
Db 1 CTGTGACCCCTTACCGCACTGTGGAAACGGAGCTGGAGAACTGACCTCTTGGGGC 60
4241 actggtctcctgacatlgaaagcagcaagcaacacag-----tc 4281
|||||
Db 61 ACTGAGGTCCTGATGTGAAAGCCAGCAGACAGCCAGCCAGTGGAGCCCTGAGCATTG 120
4282 acggttccaggtga-ttccaccccagctgctcattgatcatgatcatgttcttaagaatg 4340
|||||
Db 121 ACGGTTACACGGTGACTCCCTGCTCCAGCTGTCACTTGGATCATTTGTTAAAGTATG 180
4341 acaacattcataagccagcaactcaactcaacctcttgatta-atctggattgtgcagag 4399
|||||
Db 181 ATACTCAGA-----AGCCATACCTCACCCCTTGTTAGTCTGGGGTGTCTGCATG 234
4400 gctaagagccatgttgcctgataacctgaggagacctccacctctctgagcctataccg 4459
|||||
Db 235 GGTAGGGGACATGTTTCCAGACACC-AGTATACCTGCTGCTCTGGGCTTGCCTGACTG 293
4460 tggctctct--cactccatatggtgctgtgtcttcttgcctctcttgaaccacaa 4517
|||||
Db 294 TGGCTCTGTGACCTTCCAGATGGTGGCTTGTGCTTGTGCTTATCT---GTGGGTAG 349
4518 ctgtgacctgtggtatctctctatlgacgtlaagcacaagagggggttcaactctta 4577
|||||
Db 350 CTGTGACCTGTGGTATCTCTCTATTACTGTAGGACCAAGCAAGGCTTGTATCCCATTA 409
4578 tgtctgtatctcaatccagaaagtaactgccaacatgtgtgtcatalaataatg 4633
|||||
Db 410 TGTTCTGATTCATTAATACCAAGTACTGCCACAGGGGTGCTGCATTAATATGT 465

RESULT 15
T91781 373 bp mRNA linear EST 22-MAR-1995
LOCUS
DEFINITION yd52d09.s1 Soares fetal liver spleen 1NPLS Homo sapiens cDNA clone
IMAGE:111857 3' similar to SP:ZK783.1 CE00760 ;, mRNA sequence.
ACCESSION T91781 GI:723694
VERSION T91781.1 GI:723694
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 373)
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman
, M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J.,
Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston
, R., Williamson,A., Wohlmann,P. and Wilson,R.
The WashU-Merck EST Project
Unpublished (1995)
Contact: Wilson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
Insert size: 710
High quality sequence stops: 283 Source: IMAGE Consortium, LNL
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.lnl.gov) for further information.
Insert Length: 710 Std Error: 0.00
Seq primer: -21m13
High quality sequence stop: 283.
Location/Qualifiers
1..373
/organism="Homo sapiens"
/db_xref="GDB:467474"
/db_xref="taxon:9606"
/clone="IMAGE:111857"
/clone_id="Soares fetal liver spleen 1NPLS"

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:26:34 ; Search time 31.39 seconds
(without alignments)
4380.498 Million cell updates/sec

Title: US-09-842-930A-2
Perfect score: 7861
Sequence: 1 SLPSLLTRLEQMPDYSIFRQ.....WGHGPDMSQATTVTPR 1431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq Length: 0
Maximum DB seq Length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71.*
1: PIR1.*
2: PIR2.*
3: PIR3.*
4: PIR4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4447.5	56.6	1069	2	T22674
2	510	6.5	2871	2	A55624
3	508.5	6.5	2907	2	A57278
4	487.5	6.3	3002	2	A47221
5	496.5	6.3	1620	2	T27283
6	496.5	6.3	2918	2	A54105
7	492	6.3	2555	2	A40043
8	487	6.2	2524	2	A35844
9	484	6.2	2703	1	A24420
10	481	6.1	2871	2	A55657
11	480.5	6.1	1111	2	T26972
12	475.5	6.0	1203	2	A49175
13	474.5	6.0	2352	2	T30201
14	473.5	6.0	2437	2	SA2612
15	473	6.0	1574	2	T13954
16	473	6.0	2318	2	S45306
17	469	6.0	2321	2	S78549
18	465	5.9	2471	2	A49128
19	452	5.7	2531	2	T31070
20	446	5.7	2531	2	S18188
21	445	5.7	2531	2	A46019
22	437.5	5.6	1964	2	T09059
23	434	5.5	4135	2	T42629
24	432	5.5	3566	1	A40701
25	421	5.4	4006	2	T09070
26	413.5	5.3	1408	2	S16148
27	411.5	5.2	1722	2	B99753
28	403	5.1	1064	2	A40136
29	399	5.1	3507	2	T34513

30	392	5.0	1220	2	A56136	jagged protein pre
31	375.5	4.8	1584	2	T22674	hypothetical prote
32	373.5	4.8	1106	2	T18739	hypothetical prote
33	371	4.7	3051	2	SA2373	hypothetical prote
34	359.5	4.6	1372	2	T25933	hypothetical prote
35	357.5	4.5	1133	1	EGRT	epidermal growth f
36	355.5	4.5	3106	1	S53868	laminin alpha-2 ch
37	348.5	4.4	1786	1	MMMSB1	laminin beta-1 cha
38	348.5	4.4	2139	2	A35672	crumbs protein - f
39	348	4.4	1801	1	MMRTS	laminin beta-2 cha
40	348	4.4	4544	1	S02392	alpha-2-macroglobu
41	344	4.4	1786	1	MMHUR1	laminin beta-1 cha
42	344	4.4	3084	1	MMMSA	laminin alpha-1 ch
43	343	4.4	3075	2	S14458	laminin alpha-1 ch
44	338	4.3	3623	2	T09456	intrinsic factor-B
45	335	4.3	3672	2	T23433	hypothetical prote

ALIGNMENTS

RESULT 1
T42681
hypothetical protein DKFzp434E0321.1 - human (fragments)
C:Species: Homo sapiens (man)
C:Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T42681
R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, November 1999
A:Reference number: 222233
A:Accession: T42681
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-870:871-1069 <AA>
A:Cross-references: EMBL:AL133021
A:Experimental source: adult testis; clone DKFzp434E0321
A:Note: the cDNA sequence contains a +1 frameshift near codon 870
C:Genetics:
A:Note: DKFzp434E0321.1

Query Match	Score	56.6%	Score 4447.5	DB 2:	Length 1069;
Best Local Similarity	73.1%	Pred. No. 2.7e+268;			
Matches 798;	Conservative 100;	Mismatches 146;	Indels 47;	Gaps 6;	
QY	316	GAASCKCAAGFGNGTVCATINACETSNCGGCTKADCKRTTGNRYCVKAGYTGDTGTC	375		
DB	6	GAASCKCAAGFGNGTVCATINACETSNCGGCTKADCKRTTGNRYCVKAGYTGDTGTC	65		
QY	376	LEINPCLEMHGCGDNRNAECTGTGPNQAVNCCLPKYTGDKKVCSLINVCCLTNNGGCSPFAF	435		
DB	66	LEINPCLEMHGCGDNRNAECTGTGPNQAVNCCLPKYTGDKKVCSLINVCCLTNNGGCSPFAF	125		
QY	436	CNYTBDORICTCKPDYTGDTGIVRCGSYIGELPKNPSTQYFQLOEHAVERELAGCPPT	495		
DB	126	CNYTBDORICTCKPDYTGDTGIVRCGSYIGELPKNPSTQYFQLOEHAVERELAGCPPT	185		
QY	496	VAAPLSSSFNHEPRIKDMDQGLMSQVLRHYHVGCOOLLNLKVTTSATTTQGPVST	555		
DB	186	VAAPLSSSFNHEPRIKDMDQGLMSQVLRHYHVGCOOLLNLKVTTSATTTQGPVST	245		
QY	556	VSDPVTFINNEKAVSSDITSTNGVIVHIDKLSKRLNLTTPKDALGRVLOLVTVAAHH	615		
DB	246	VSDPVTFINNEKAVSSDITSTNGVIVHIDKLSKRLNLTTPKDALGRVLOLVTVAAHH	305		
QY	616	GVTFRSKLIDSGSLSVITDSIHPTVTFWPTDKALALPPOQDFLPMQDKDKLKYTL	675		
DB	306	GVTFRSKLIDSGSLSVITDSIHPTVTFWPTDKALALPPOQDFLPMQDKDKLKYTL	365		
QY	676	KHHVTRDSKALASDLPKRSASMTTLOGSELISVRCGSDIGELFLAEOMCRFTHRGLEPDV	735		
DB	366	KHHVTRDSKALASDLPKRSASMTTLOGSELISVRCGSDIGELFLAEOMCRFTHRGLEPDV	425		

QY 726 GVAIVDCLLNNPRLTGSGCDFTTFEDIGEGCGSCITFFPKCPKLSPKRVKKCKLYTNPLPF 795
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 426 GVAIVDCLLIDLPILGGRKDPTFTFDASGECGSVNTPSCPWMSFPKVGVKOKCLYN-LPF 484
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 796 RKNVECCNLCTVVIQTFRCCGHGYFMPDCCOACPGVPDPNCRNRGMCRDLVPRMGQLCHT 855
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 485 KRNLEGGCERERSLVTIQIRPCKKGYYGRDCCOAPGPGDPAPCNNRGVCLDLOYSATGECKCNT 544
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 856 GFNGIACELCMHNGRGPDCQPRSCSEHGQCDBGITGSSBCLCEGTGWTAASCDTPFAVAVAY 915
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 545 GFNGIACELCMCWGRGPPDCLPCGCSBDHGQCDGDTGSGQQCLCEGTGWTPSDDTQAVALAVAY 604
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 916 CTPACSVATCATENNTVCVNLTNEYGDGITCYVVDPCFNKNNNGCAKAVAKCSQKGTQVSCSC 975
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 605 CTPPCSAAATCATKENNTCCNELDYEBDGITTCYVVDPCRKDNNGCAKAVAKCSQKGTIVSCSC 664
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 976 KKGIYGDIQSCTEIDPCADGVNNGCGCHEHATCRMTRPGKHKECKSKSHYVGDVDEPEOLP 1035
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 665 QKGYGDSHCHEIDIPCADGLNGCGCHEHATCRMTRPGKHKECKSKSHYVGDDILNEPEOLDP 724
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY 1036 IDRCIDLDNGGOCPPRDASCADLYEFODTTVGPHLRSPFLGQYLTFPEKARECAKAEATITY 1095
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 725 IDRCIDLDNGGOCCHADKACYDLHFODTTVGPHLRSPFLGQYKLTPEKARACANEATIMATY 784
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1096 NOLSTYAOKAKYHLCSAGMLIESGRVAYPTTYVASQKCANVAVGIVDYSRANKSEMWDVFCY 1155
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 785 NOLSTYAOKAKYHLCSAGMLETGRVAYPTTFASQNCGSGVGIIVDGPBRNKSEMDVFCY 844
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1156 RKKDVCNCFCKAGYVDGRSGCSGNLLQVLMSFSLNPLTEVLAFSKSSARKQAFLKHLTD 1215
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 845 RMKG-----SAGLFQQLSRPCIS-----RPDD 868
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1216 LSIKGTLEVPONSGLLPGKKSLSGRDIEHHLTNVANSFYNDLVNGTFLMTLMSQLITFS 1275
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 869 LSIKGTLEVPONSGGENETTLGRDIEHHLANVSMEFINDLVNGTTLDTRLGSKLLITAS 928
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1276 ODOLH-OETRFPDGSIILOMDIIANGILTHIISEPLRAPPTAATPAHSGSLGTGFCAVYL 1334
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 929 ODPLOPETREYVDGAILLQMDIFASNGLIHVISRLKPAPPAVYLTLRHGLDAGTFPAILL 988
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1335 VTGAIALAAVSYFRLKQRTFGRFDOKRRTLMWSLAS-----SSPRIISOTLCMRPQR 1388
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 989 VTGAVALAAVSYFERINRRITIGQHFESEEDINVALGKQEPENISNPLEYESTTSAPPE- 1046
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 1389 HQQSPVPTPSQ 1399
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1047 -PSYPDFTDSE 1056
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 2
A55624
fibrillin-1 precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 23-Mar-1995 #sequence_revision 23-Mar-1995 #text_change 11-Jan-2000
C:Accession: A55624
R:Yin, W.; Smiley, E.; Germiller, J.; Sangulneti, C.; Lawton, T.; Pereira, L.; Ramirez,
J. Biol. Chem. 270, 1798-1806, 1995
A:title: Primary structure and developmental expression of Fbn-1, the mouse fibrillin ge
A:Reference number: A55624; MUID:95130561
A:Accession: A55624
A>Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <YIN>
C:cross-references: GB:I29454; NID:g575509; PIDN:A5A56840.1; PID:g575510
C:genetics:
A:Gene: Fbn 1
C:superfamily: unassigned EGF-related proteins; EGF homology
#:I201-I236/Domain: EGF homology <EGF>

Query Match	6.5%	Score 510;	DB 2;	Length 2871;
Best Local Similarly	21.2%;	Pred. No. 3.7e-23;		
Matches 311; Conservative	130;	Mismatches 416;	Indels 612;	Gaps 85;

QY	117	ATDKGVHGLEKYLEIÖKNC	-----DNNDTIYR-GEC-----GKCSQAP--	157
Db	293	STIPGVCGGECTTAVSSYCKCPGPGHYTSPDGTICVDVPRPGCYATALANGSCSNLPQS	158	-----CPL-----ETKPLPER--RKCTISYEMK-----180
QY	353	ITKMOCCDDLGRMSPGVYAPAEPCPTKSTEDPKNKCSPLVLPGRPREPPRPPGRLPPV	161	-----RSVEIGCQPCVFTI--ITRACWLAS-LAHNAK--PAPG--214
Db	413	QVPVPGPGGVILVPRPRPELYLPSREPRVLPAVNTVYCOLVRLCQNGRCIPTPSGX	215	-----EVMKCALGIASVWDVNGNG--TCQOIGGFNPTAECTBEKY255
QY	473	RCEKNKGÖLDINGECIDVBECKEKNPCTGEGCINNOSTYCHCRAGYOSTL--ITTECRD	530	530
Db	256	GIHCDQ--ACSCVHRCSSQGLPGDS--CDPDVW-----RGVNC-DM-ETTDN-C-NGTC	304	304
QY	531	IDECIÖNGRICINNGRCIN--TDGSPHCVCNAGFAHVSSGKNCEMDDECRFPNMGPNMG	587	587
Db	305	HTSANCLDPPGKAKSCCAAGFR--GNGIVCTA INACET-----SNGC-----	346	346
QY	588	-----INEDGSFKCKICPGFOLASDGRYKXDINCEPEPGICMNGRCVNTDGSYRCECF	640	640
Db	347	-----STRKAD-----	352	352
QY	641	PLAVGLDGRMCVDTIHRSSCYGGRGCQVRLPFAVYKSECCASFTYAFGEPCQCP	700	700
Db	353	-----KRTTPGNRY-----CYKAGYTG-DG372	372	372
QY	701	AÖNSAEYALCSPGPGTISAGTIDINECALDPDICPGNIGENLRTGYKICNSGVEVDITG	760	760
Db	373	IYCLEINPCLENHGCGCRNAECTGTGPNQAVNCJLPK--YTGDGVCSLIINWCLTN--	426	426
QY	761	KNCVDINCVAINSLDCL--NQGCRT--PGSYVCTC--PKGVYKPRDLKTECEDIDECSSPCI	817	817
Db	427	NGGC--SPFAFCNTBEDÖRICTCKPDYTG-D--GIYCRSIIYELPKPNBSTÖYFFÖLÖE	482	482
QY	818	NGVCNKSNGSP-----ICBSPESITLDPKTICITETIKGT-----	852	852
Db	483	HAVRELAPGPGFTYFAPRLSSFNHPRIKMÖDQGLMSÖVLRYHVYGOÖLLD--NLKVT	541	541
QY	853	-----CMÖVYIDRCETIN865	865	865
Db	542	TSATTLQ-----GEPVSI-----SVSÖDVF INNEAKVLLSDI	574	574
QY	866	INGATLKECSSLSGAAMGSPCTICQDLPICGKGFRIKGTÖCEDI--NECEVFPG--	919	919
Db	575	ISTNGVI--HVIDLKLSPRNULITRKDALGRVLQMLTVAANHGYTFKSLIQDSGLLS	631	631
QY	920	VCKNGLCVNRSFKCEPCNGMTL--DATGRICIDIRU-----ETCFLYKDEECTLP	970	970
Db	632	VI-----FDSIHTPVYVWPTD-----KALEALPREÖDPLEFONDKNCKLSYLKF	677	677
QY	971	IAGRHRMADCCCSGAMMGHEBECEPLNRSRYEELCRGGAFA--TADITNGKRF	1026	1026
Db	678	HVIRDSKALASDLPRASMKTÖGSELVRCGTGSDIGELFLINÖMCRFHIGLLEIDGV	737	737
QY	1027	KDINECKMIPILCHGKCRMTI--GSFKCKDSGFAIDS--EERNCT-----	1069	1069
Db	738	AVGIDCLMNPÖTGGRCDDTFTFDIGEGSCSFLTP--KCPLKSPKG--YKKRCIY	790	790
QY	1070	--DIDECRISFDLGR-----GÖCVNTPGDEPECCD--EGYESGMMKKNM--	1112	1112
Db	791	NPLFRNRVCCQMLCVIVIQTPRCHGYEMP-----CQACPGPGPTPCPNRMRCDL	844	844
QY	1113	--DIBECQ-----RDLRLRGGICHINTBSTGYCECPRGHÖLSP--NISACIDI	1156	1156
Db	845	-----YTPMG-----QCLCHTENGNT-----ACE-LCWA--	867	867
QY	1157	NECELSANLCPHGRCVNLICKYÖCAGCPGHNPTHRILFCVDDIDECSSIMNGGCEFTTCTUSD	1216	1216

Query Match	6.5%	Score 508.5	DB 2	Length 2907
Best Local Similarity	23.5%	Pred. No. 4.7e-23		
Matches 296	Conservative 123	Mismatches 395	Indels 445	Gaps 87
QY	190	OCVRIITFRACLAHLASLHAKPAPG--EYKMCALGTAVMGVNGTGR--CQCGLGFG	244	
DB	502	KCLPTVSSSTR-CRNMGYK-QDANGCICDIVDECTSNPCSGCVNTPMGSYKCHAGFOR	559	
QY	245	T-----ACETCTEGKYGIHEDQ-ACSCVHRCGSGPLGDS--CDCGVGRVKCDMEIT	297	
DB	560	TTTKQACIDIDE-----CIQNGVLCKNRGCVN---SDGSFQICNAGF-----ELTT	603	
QY	298	D--NC--NGCTHTSANCL-----LDPDKASCKCAGF--RGNGTVQTAINACT-----SN	343	
DB	604	DGNKVCDHDECTTTNMCLGMCINEDGSKFYCKRPFILAPNGRCCTDVDEQOTPEICNN	663	
QY	344	GGC-----STKADCKRTTPG-----NRVCV-----	363	
DB	664	GHCINNEGFFRCDC--PPGLAVGVDRVCVDTHMRSTCYCEIKKGVCPFPFGAATKSE	720	
QY	364	-----	363	
DB	721	CCCANPDVGFGEPCPCPAKNSAEFHGLCSGIGITVDGRDINECALDPICANGICENL	780	

RESULT 4
A47221
fibrillin 1 precursor - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 02-jun-1995 #sequence_revision 25-Apr-1997 #text_change 21-jul-2000
C:Accession: A47221: 154355: S17064: 155574: S62111: A34198
R:Cosmic: G.M.: Chalberg, S.C.: Dietz, H.C.: Charbonneau, N.L.: Sakai, L.Y.
Genomics 17, 476-484, 1993


```
Db 1573 KACE-----DIDECISLPNICVFCTCH-----NLPGIFRCECEIG-YEL--DRSG 1613
QY 1083 EAC 1085
Db 1614 GNC 1616

RESULT 5
727283
hypothetical protein Y64G10A.f - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27283
R:Alnscou9n, R.
submitted to the EMBL Data Library, September 1999
A:Reference number: Z20356
A:Accession: T27283
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1620 <WTL>
A:Cross-references: EMBL:AL110498; NID:e1542303; PIDN:CAB5471.1; CESP:Y64G10A.f
A:Experimental source: clone Y64G10A
C:Genetics:
A:Gene: CESP:Y64G10A.f
A:Introns: 77/1; 116/1; 198/1; 282/1; 365/1; 425/1; 466/1; 548/1; 559/1; 601/1; 625/1; 7

Query Match 6.3%; Score 496.5; DB 2: Length 1620;
Best Local Similarity 20.3%; Pred. No. 1.3e-22;
Matches 294; Conservative 126; Mismatches 403; Indels 625; Gaps 83;

QY 129 VLEIQK-----RCNDNDIIVAGEGCKSQQAPCL---ETPRLAETKRCITSTFMGKR 181
Db 72 LIRVANCASADLCIHNGT-CVPSEHNDNEQVCEPVGFTGAKQYDANEC-----MANN 124
QY 182 SVETICQPCQCVRTIITR--ACWLASLAHNAKPAPG-----EYMCALGTASYWD 228
Db 125 G---GCEHCVCVNTIGTYCRGW-----PGFELSGDQNTGSDIDECVANGGSD 170
QY 229 G-VNGTG-----TCQCGLG 242
Db 171 RCVNSPFGFCRDCPSDLYLHADGRTGKVTSCSTDNCGCEHCENDSNGEYRCRGRVF 230
QY 243 -----NGTAC---ETCTEKGYGICHDQACSCVHGRCSSQGPLGSGSDCVGMGVACDME- 294
Db 231 KLSNKRSCQPYVPCFDNKG--COHCTNNHGR-----AQCQCYPGFH-LSYDRRS 279
QY 295 -ITTDNC---NOTCHTSANCLIDPGKASCKCAAGFR--GNGTVG-TA1INACETSNCGCS 347
Db 280 CVDIDECAKNNGCHEFCENV---KGTFRCKCREGYOLGRDRTCEEMLGCGOVANGGC- 334
QY 348 TKADCKRTTPGRCVVCAGY----- 368
Db 335 -OHDCYDQPDGHNVCCKNGYILANDOKLCHDNISTVILHARAPRLDSYETVTCVTPDL 393
QY 369 -----TG-----DGIVCLEINPCLENHGGCDNACTQGRQA--- 402
Db 394 TCHKICMLHDSGHVOCFDDGDELIDSKFCODINECHENN---DCSQICVNLASGV 447
QY 403 VCNCLP--KYTGDKVSLINVCILTNNGSGSPFAFCNTEDOPRICTCKPDV--TGDCIV 458
Db 448 ECQCGPGRFLMKDKRTCDIDSECSNNGCF--QICS-NQEGGYMCSCEPGELESDGHS 504
QY 459 CRGSITGELPKNPSTSYFQLOEHAHVR-LAGPGPFVFAPISSFNHPRIKWDQOG 517
Db 505 C-----HDMNECLINNG----- 516
QY 518 LMSQVLRHYVVGCOOLLDNILKVTLSATTLQGEPRYSISQDTPIINNEAKVLSDDITST 577
Db 517 -----GCAO-LCKNRGSRKCCFAGY-ILAHDEKSCVAASADIFESDIDEDY 563
QY 578 NGV--IHVIDKLSPKNLLITPKDAIGRVLONLITVAANHGYTFKSKLIDSGLLSVITD 635
```

```
Db 564 SKVPGIDSIDELYS--SLESYPADSEPRFL-----VFGRRRHVAKCVNPGTSLSELS 616
QY 636 SIHTPTVTFWPIPDKALBELPREQDFLEFNQDNKKDLKSTLKHVIRDSKALASDLPRSA 695
Db 617 EVRTD-----PSEKCPNGF-----FGSITCOLSCSDCQNG 646
QY 696 WKTLOGSELVRCGTGSD-IGELFLNEMQCRFIHRGLLFQVGVAGIGICLLMNPPLGG-R 753
Db 647 KCSMRGSLSKDCDPSGTYGEC--EOTCRNGYWG---YDCAHKSCKLCDDPSTGSCR 700
QY 754 CDFTFTFDLPGRG--GSC---LFTPKPLKSK-----PKGVKKKCIYN 791
Db 701 CBD-----PEKCSGPRCDPGRFGYSQCNLCKCMDEPRNGRCDPVGCTCPDLYQSGCEK 754
QY 792 PLPF-----RRNVEGQNL--CTVVIQTPRCCHGYFMPDQO----- 825
Db 755 PCPHFTFGKNCNCFPCKCARENSEGCDELITGKC-----CKPYGYGHNCKRMCSPLFG 807
QY 826 -----ACPGS-----PDTF-----CNNRGMCRDLITP 847
Db 808 ACCAMKSCSPAGIRCDPVTGDCCTKKCPAGYQGNLCDDQCPAGYFGYDCEQKSCADVASP 867
QY 848 M----- 848
Db 868 HSKYCHHVTGCTCLPKGTGPLCDQCLFYETIEFDIAFSINYACAPNTYGPNCATHC 927
QY 849 -----GQCLHTGFNGTAC-ELCWHGRGPRDC-QPRSCSEHQGDEGITSGE 894
Db 928 SCVNAKCDSEDSGCHCTPGFYGATCSFVCPGRFGIDOMOLCKONGALCD---TSNGS 984
QY 895 CLCETGTAAASDTPTAFAVCTPA-----CSVAHTCF-----NNTCYNLNIEGDG 942
Db 965 CECAPGWSGKKCD-----KACAPGTFGKDCSKKCDKADGMHCDPSDECTCPGKKGHK 1038
QY 943 I--TCTVVDPCQNNKGCAKVAKCSQKGT---QVSCCKKGYKGDGYSCEIIPCACGV 996
Db 1039 CDFTDSCGF--GAGCKGICSCGNATGDSVYSCSCREGRMRK--KCDR--PCPDGR 1090
QY 997 NG-GCH-----EHAT--CR---MTGP-----GKHK----- 1015
Db 1091 FGEQGNALICDCTTNDTSMYNPFAVARCHVTEGECRCAPAGMTGPRDQCTSLGRHSEGRH 1150
QY 1016 -----CECKSHYVDGYDCE-PEQLPLDRLCQ-----DNQCGH--DA 1050
Db 1151 SCQGSNGASCDRVYTGFCDCPSGFEMKNCSESECPGLWGSNCMKILCLMHGECNNENGD 1210
QY 1051 SCADLYFQDPTVGVFHLRSPLGOYKLTFDKAKAEACAKEAATATATYNQLSYQAQAKYH 1110
Db 1211 ECIDGWTGPSL-----CPTQPF-----GRNCAQRC---NCKNGASCDRKTGRCECL 1253
QY 1111 AGWLESGRAYVPTTYASQKCGANVGVIVDYSRANKSEMDVECYRMKDVNCTCKAGYVG 1170
Db 1254 PGM--SGE-----HCEKSC--VSG--HYGAKCBET-----CECENALC 1286
QY 1171 DGFS--CS 1176
Db 1287 DPISGHCS 1294

RESULT 6
fibrillin-2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 09-Sep-1994 #sequence_revision 09-Sep-1994 #text_change 17-Nov-2000
C:Accession: A54105; S17063; S31101
R:Zhang, H.; Apfelroth, S.D.; Hu, W.; Davis, E.C.; Sanguinetti, C.; Bonadio, J.; Mecha
J. Cell Biol. 124, 855-863, 1994
A:Title: Structure and expression of fibrillin-2, a novel microfibrillar component pr
A:Reference number: A54105; MUID:94165150
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual
A:Molecule type: mRNA
```


QY	375	CLINPCLNHHGCG-DR-----	NACCTQTGP--NOAVON-----	405
Db	530	C-QSNPCL-MDGTCHDINGFKSCALGFTGARCOIINDIDCCOQSPCRNRGICHDISIAGS	587	
QY	406	--CLPKRYTGDKKVCST-LINVCLTN---NGGCSPEAFCAVTEODORTCTCKPPTYDGIYC	459	
Db	588	CECPRGYTGTS--CETININDCSNPPCHRGK-----IDVANSFKCLCDPGYT--GYIC	636	
QY	460	RGSITVELPKNP-----STSQYFPOLOEHAVBELAGPPTVFAPLSSFNHHPRI	510	
Db	637	OKOI-NECESNPCQFDHCODRVSYYCCQO-----AG-----TSKNCNEVNV	678	
QY	511	KDMOQOGLMSQVLRYHYVGOQOULLDNLKVTYSATTLQGEPPSISVSQDTVINNEAKVL	570	
Db	679	NE-----CHSNPC-----NGGAT--	691	
QY	571	SSDIITNGVIVHVIDKLSPKNNLI--PKDALGRVLONLTPVAAHHGTFRSKLIQDS	627	
Db	692	-----CIDGINSTKQOCVGRGFTGQHKEKVNDECISPPCANNG-----	728	
QY	628	GLLSVTITDSIHTTPVYFWPPTDKALEALPREQODFLFNQDNKDKLKSYLKFHVIRPSKALA	687	
Db	729	-----VCIDQV-----NGYKCECPRGFYAHLSDVDECA	758	
QY	688	SDLPKRSAMKTLQG-SFLSTRCGTG-----SDIOELFLNEMQCFHRLGLFVGVAY	739	
Db	759	SN-PCVNEGREDDEINFICHPRGYTGKRCLEIDECSSNP--CQ--HGCTCYDKLAF	813	
QY	740	GIDPLLNPTLGGRCDFPFTFDIPGEC--GSCI-----FPPK-CLPKSKPK	782	
Db	814	SCQC--MPGYTGQKCEFINIDCVTNFCQNGGTGIDKVNQKCVCKVCPFGRCESKMDPC	871	
QY	783	GVKR-----KCI--YNELPR-----RVNEGCONL-----	805	
Db	872	ASNCKNAEACTPSSNFLDSPTCKLGYTRYDEDDIDBCSLSSPCRNACSLANVPGSRY	931	
QY	806	CTTVIYIOTPRCHGTFYMDC---QACBGDPFPCCNRGKRCMLYTPBMG--QCCLHTGFG	859	
Db	932	CL-----CTKGEYGRDCAINTDDC--ASFPCCONGGTCLD--GIDGYSLCYVDGFGD	978	
QY	860	TACEL-----CWHGRF-----GPDQCPRECSRHQOD	886	
Db	979	KHCETDINECLSOPCOGATCSQYVNSYTCCTPLGFSGINCOTNDEDTRESSCLNGSGCI	1038	
QY	887	EGITGSEGCLETGWTAAASCDPTPAFAVC--PPACSVHAHTCIENN---TCVNLWEYEDG	942	
Db	1039	DGINGY-NGSGLAGYSANQOYK--LNMCDNSMPLCATGCHQONNEYTCHPSFTGK-	1093	
QY	943	ITTC--VYDFPKONNGGCACAKYAKCSQKQTOVSCSKKGYGDEYSC-TEIDPCADV---	996	
Db	1094	-QCSEYVDMDGOS--PEENATCSOMKHOPSCSAGMTGK--LCDVOTISCODAADRK	1148	
QY	997	-----NGGCHHAHCMTGTGKHKCEKSKHYVDGVDCEPEQPLDRLCLQDNGQCHP	1048	
Db	1149	LSLRQLCNNGCKKQYK-----NSHWYCSQGYAS--YCKE--IDEC--OSOPCON	1194	
QY	1049	DASGADLYFODTTPYGVFHLSPGLQYKLFEDKAKEACAKEAATIAITYNOLSYAKAKYHL	1108	
Db	1195	GGTGRDL-----IGAECCRCRGFGQGNQNELINDCAENP-----	1228	
QY	1109	CSAGMLESGRVAYPTTYASQKCGANVGIADVGRANKSEMDVFCY-----RMKDVN	1161	
Db	1230	CONGCTCHDRV---NMFSQSCPPTGMIIT---CEINKDCKPKACHNNGSCIDRVGFE	1283	
QY	1162	CTCKAGYVGDFGFSGNLLQVL	1183	
Db	1283	CVCPGFV--GARCEGDINECL	1302	

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 11-Jan-2000
C:Accession: A55567
R:Illstra, D.J.; Li, L.; Potter, K.A.; Womack, J.; Byers, P.H.
Genomics 23, 480-485, 1994
A:Title: Sequence of the coding region of the bovine fibrillin cDNA and localization
A:Reference number: A55567; MUID:95137597
A:Accession: A55567
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-2871 <TIL>
A:Cross-references: GB:L28748; NID:9508427; PIDN:AA414122.1; PID:9508428
A:Superfamily: unassigned EGF-related proteins; EGF homology
F1201-1236/Domain:EGF homology <EGF>

Query Match	6.18;	Score 481;	DB 2;	length 2871;
Best Local Similarity	20.88;	Pred. No. 2,4e-21;		
Matches 276;	Conservative 128;	Mismatches 383;	Indels 542;	Gaps
QY 125 GLEKLELQKNRCC-----NNDITIVRG-----ECCGKSQAQCPLETPRLRET 168				
Db 1100 GYSESEFMKKCMIDECQDPRLCRGVCLNTEGSRCECPGHOLAP---NISACIDI 1155ACIDI				
QY 169 KRCIYIYFMGRKSRVFTGC-QPOCVRTIITRAOCLASLAHNAKPAEGVKMCAALGAYW 227				
Db 1157 NECELSAHL-----CPHGRCVNLICKYQACANGVHNS---TPDRFLVCVIDECSIM 1200				
QY 228 DG-----VNGTgr-CQCGIGF-----NGTACE--TCFE--GKYGTH 258				
Db 1205 NGGETECTINSEGSYECSCQPGFALMPDRQSCDIDECEDNPMTICGGQCTTNPGEYRCL 1266				
QY 259 CDQA-----CS-----CVHRCGSGPLGDSGCCDYVMGRGVK-----C- 291				
Db 1265 CYDGFEMASEDMKTCVVDVNECDLNPNICLSGTC-ENTKGSFTICHDMDGIGSKKGTCTDID 1322				
QY 292 -DMEITTDNC--NGTCHTSANCLIDPDGKASCKAAGFRNGTAVTCAIACETSNMGCGST 348				
Db 1324 NECEIGANDCRHNAVINTTA-----GSPKSGCSQPMIGDGIKCTIDIDECSSNGTHMCSQ 1376				
QY 349 KADCKRTTPGRKVCVCAAGTYIGGIVCLLEINPCLLENHGGCDRAECTOTGPMOAVNC-- 406				
Db 1377 HADCKNTM-GSYRCLCKEGYTGIGGFTCTDIDECSENMLDGG-NGOCL-NAPGGYRCECDM 1433				
QY 407 -LFRYTDGKV-----GSLINVC-----TNNGCGSPFAPC 436				
Db 1434 GFVP--SADGKACEDIDECSPINVCYFGTCHNLPGLFRCECEIYELDRSGGNCIDVNEC 1491				
QY 437 -----NYTEDDQRICTCKPDY-----TG 454				
Db 1492 LDPTTCISGNCVMTPGSYT-----CQCPDPFELNPTRVGCVDTRSGNCYLDIRPBGNG 1546				
QY 455 D-----GIVCR-GSYIG--ELPRNPSTSYTFOLOEHAVRELAGR----- 492				
Db 1546 DTACSNELIGVSKASCCSGISGKMGWGPCLCPRVMTSEY-----KILCPGEGF 1595				
QY 493 ---PFTVAPLSSFNEHPRIKIMDOOGLMSQYLRYHVVSCQOL--LIDNMLVVTTSATPL 547				
Db 1556 RPNPTTVI-----LEDIDE-----COELPRLCOGKGCINTFGSF 1622				
QY 548 QGE-PVSTISVSQDTVF-----INNEAKVLSSDIISTNGVLIHVD 585				
Db 1630 QCRCPGTGYLINEDTRVCCDDVNECEPTGICGPGTCYMTNGVYTCICPPDYMYQNGGNCMD 1689				
QY 586 --KILSPKNLLITPKALGRYLQNLTTVAANHGYTFKSKLIDSGLLSYITDSIHPTVY 643				
Db 1690 MKRSYLCARNYADNOTDGELELNMTKKMKCCSYN-----IGR 1722				
QY 644 FMPYDKALEALPRPQOQFLFNQDMKDKLYLKFHYITRDSKALASDLPKRSAMKTLQSGE 703				
Db 1728 AM-NKPEQCP-----IPSTDEFATLCGSO 1755				
QY 704 -----LSVRCGTGSDIGELFLFNQDMCRFTIHRGLLPDVGAAYGIDCLLMPMTLGRCDPFT 758				

```

Db 1752 RPEGVLDIYGLPVDIDE-----CR-----1771
QY 759 TFDPEGC--GSCI-----FTPKCPILSKPRGVAKKCIYN-PLPFRNNVEGCQN--LCTV 808
Db 1772 --ELPGVCNVCINNMVGFRCPCPV-----GFFYNKLKLVCDIDBCQNGPVCQ- 1819
QY 809 VIQPRCCCH--GYRPPDCQACPGSPDT--PCNNRGMCRDLYTMMGCLCHTGNNGTACE 863
Db 1820 --RNAECINMAGSYRCDC--PGYRFTSTGQCNDNRNECOEI-----1856
QY 864 LCMHGRGPPDCQPRSCSEHGQCDEGIISSGECILETGTATASCDLPFAVAVCTPACSVH 923
Db 1857 -----PMICS--HGQCIDTV--GSFYCLCHTGTNADOTMCLIDNECERDAGN 1902
QY 924 ATCTENNT-----CVCN-----LNYEGDITCTVDFCKONNGCAKAVAKSOKGTQVSCS 974
Db 1903 GTC--RNTISGFNRCNNGFLSHINN--CIDVDECATGNLNCRNQCINTVGSFQCQ 1957
QY 975 CKRGYK--GDGYSCIEI-----DP--CADGVNCGCHEHATRMTPGPKHKECKSHIYGD 1025
Db 1958 CNEGEVAPDGRFTCVDIENECLIDPRKCAPG-----TCQNL--DGSYRCICPPGYSIQ 2007
QY 1026 GVDEPEPLDRLCLQDNGOCHPDASCADLYFQDTTGVFHLRSPILGQYKLPFKAKKAC 1085
Db 2008 NDKE--DIDEEVEPEIC--ALGTCSN-----TESFKCLCPDG--FSLSTGRK 2053
QY 1086 AKEAATATVYQNLQSYAKAKYHLCSAGWLESGRVAYPTT--YASQKCGCANVVG-----1136
Db 2054 -----QDLRHSYCYAKF--EGKCKSPKSRHNSKECCCALKGEWGDPC 2096
QY 1137 -----IVDYS-----RANKSEMDVFCYRNMKDYN-----CTCKA 1166
Db 2097 ELCPTEPEAFQIOLPYSGIIVGPDSDAVDMDECKEPDV--CKHGQCINTGDSYRCPCP 2155
QY 1167 GYVGGGFCSC 1175
Db 2156 GYILOGNEC 2164

RESULT 11
T26972
hypothetical protein Y47H9C.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: T26972
R:Harris, B.
submitted to the EMBL Data Library, October 1998
A:Reference number: Z20293
A:Accession: T26972
A:Status: Preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1111 <FILL>
A:Cross-references: EMBL:AL032657; PIDN:CAA21739.1; GSPDB:GN00019; CESP:Y47H9C.4
A:Experimental source: clone Y47H9C
C:Genetics:
A:Gene: CESP:Y47H9C.4
A:Map position: 1
A:Introns: 50/7; 84/2; 150/1; 238/3; 342/3; 797/1; 851/1; 947/2; 1017/1; 1083/1
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology

Query Match 6.1%; Score 480.5; DB 2; Length 1111;
Best Local Similarity 22.2%; Pred. No. 7.9e-22;
Matches 251; Conservative 98; Mismatches 341; Indels 441; Gaps 70;

QY 115 NVAIDKGVV--HGLEKYLEI-----QKNRCDN-----NDITIRGEGCKGSQAQPCLETR 163
Db 43 HVCVTYKTIYDDLEKKVHFTVYVNDTEQCLNPLTGFCQTEVERGO--KASYOROL--VK 97
QY 164 PLREPRKCIYSIYFMGRKRSVFGCOPQV--RTIITRACWLASIAHNAAPAGEVMCA- 220
Db 98 KEKYKQCCDGYQTKDHFCLPDCNPPCKKCKIERGKC-----ECDPGYGG--KYCAS 149

```

```

QY 221 -----LGNASVWDVGNCT-----GTCCGGLGFGNGTACE--TCTEGKXYGHCDQASC 265
Db 150 SCVSYGTMLGCKSKSCDDEGNANDPELGTCTISGFGERCERKCPDNKMGPNVYKSPC 209
QY 266 VH--GRCSQGPLGDSGCDQDVGMR-----GVKCDMEITLTDNCNCTHTSANCL 311
Db 210 QNKGCKNK-----BCKVCSDSGWGEFCLNKCEBKFGAEKFE-----CN--CONGATC- 257
QY 312 LDPDGKASCK-----CAAGFRMGTV--CTAIN--ACETSNQ-----344
Db 258 DNTNGKCIACKSGYHAGALCENECVGFSGCTQKCDCLNNOCMSSSSECKCIQWTGKHC 317
QY 345 --GGST--KADCKR--TTPG-----NRVCVCAGYTGDQIVCLEINPCLEN 384
Db 318 DIGSGRGRFGLQCKONTCPLEFSDSNASCDATGQCQCESGKGP--KDEKRCDAEQ 375
QY 385 HGG--CDRNAECTQ-----TGPNOAVCNCLPKYTGDKVCSLINVCLNNNG--CSPAFC 436
Db 376 YGADCSKTCICVRENTLMCAPNTGFCRCCKPFGYGN--CEL--ACSKDSYSPNCEKQAMC 431
QY 437 NYTEDQ-----RICTCKPDYTGDIYVGRGSIYGELEPRNPSTSOYFPDLOEHAVRELAP 491
Db 432 DMNNASECNPEPTEGSCVCKPGRGT-----KNCSE-----P 460
QY 492 GPTFYFAPILSSFPNHEPRIKDMDOGLMSQVLRVHVGCQQLLDLNLKVTTSATTLQEP 551
Db 461 CLPDRYGP--NCAHQCOC--NORG-----VCD-----484
QY 552 VSIYSQDTVFTEINAEKVLSSDIISTNGVIHYIDKLSPKMLLTPKDALGRVLQNLTV 611
Db 485 -----GADGKCOCODRGWGTGH 499
QY 612 AANHGYTFESKLIQDSGLSYTDSIHFPVYVFWPTDKALEALPEQDFLENDNNKML 671
Db 500 RCEH-----HCPADTF-----510
QY 672 KSYLKEFHIRDSKALASDLPBSASWKTLOGSELASYCGSGSDIG--ELFLNEQMRFIHRG 730
Db 511 -----GANCEKRCCKPRKIGCDPITGECTCPAGILOG 541
QY 731 LLEFVGV--AVGIDCLLMNPPLGRCDTFTTFDIPGEGSC--ITPKCPILK--SKPRG 783
Db 542 ANCDIGCEPGEBSYGPCKLHCKCVNGCKOKET-----GEC--TCQPGFFGSDSCTTCSGK- 594
QY 784 VKKKCIYNLPERRRNVGQCQNLCT--TVIQTGRCHGYRMPQCQACPGSP-----DT 833
Db 595 -----YG-----ESCELSQPCSDASCSKQTKC-----LCPLGKGVSCDQ 630
QY 834 PC--NNRG--MCRDLYTPM-----GOCL--CHTFGNGTACE--LCMHGRFGPDCQ--RS 878
Db 631 KCDPNTFGLCOEYVTPSPSCASTDEKNGVCLSPRGSSGITHENHCPSAGSYDGDQOYCS 690
QY 879 CSEHQCDEGIISSGECILETGTATASCD--TPYAVF-----AVCPACSVNATCTE--NNT 931
Db 691 CADGHGCD--PTTGECICEPEYHGKTCSEKCPDKYGGYGCALDPCPKASGSTCHINGL 747
QY 932 CVCNLTNEBDGITTCYVDFPCRKONNGGCAKAVAKSOKGTQVSCS-----CSCKKRYKKDGYSC 986
Db 748 CICPAGLE--GALCTRPSCSAGFWGNGCRQVCRCTSEYKQCNQTBECSCPAFGQD--RC 803
QY 987 IELDPCADGVNG--GCHENHATCRMTGPG-----KHKCECKSHVYGD-----GVDC 1029
Db 804 DK--PCEDGYTPDCTIKKCKCGATSSCNRYSGACHHPGTGEGCHALCESTFGLKC 861
QY 1030 EPEOLPDRCLQDNGOCHPDASCADLYFQDTTGVGFHL--RSPILGQYKLTED 1079
Db 862 SKE-----C--PKDGGDGYEEDDAIIGCHVDMSCGKAKQKFE 898

RESULT 12
A49175
Motch B protein - mouse (fragment)
N:Alternate names: Notch homolog

```


C:Species: Mus musculus (house mouse)
 C:Date: 21-Jan-1994 #sequence_revision 05-Jan-1996 #text_change 20-Sep-1999
 C:Accession: A49175; PH570; S32113
 R:Lardelli, M.; Lendahl, U
 Exp. Cell Res. 204, 364-372, 1993
 A:Title: Motch A and Motch B--two mouse Notch homologues coexpressed in a wide variety of tissues
 A:Reference number: A49175; MUID:93178563
 A:Accession: A49175
 A:Status: preliminary; nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-1203 (LARD>
 A:Cross-references: EMBL:X68279; NID:9287989; PIDN:CAA48340.1; PID:9287990
 A:Experimental source: embryo
 C:Comment: This protein has many EGF repeats and 11n-12/Notch repeats.
 C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
 F:143-174/Domain: EGF homology <EGF1>
 F:482-513/Domain: EGF homology <EGF1>
 F:560-591/Domain: EGF homology <EGF>
 F:674-705/Domain: EGF homology <EGX2>
 F:712-743/Domain: EGF homology <EGF3>
 F:836-867/Domain: EGF homology <EGX3>

Query Match 6.0%; Score 475.5; DB 2; Length 1203;

Best Local Similarity 23.0%; Pred. No. 1.8e-21; Matches 256; Conservative 112; Mismatches 412; Indels 341; Gaps 70;

QY 219 CALGTASWDGVNCTGTCCGCLGNGTACETCTEGKYGIHCDACS--CVH--GRCSOGPL 275
 DB 67 CHKALCDTNPRLNQLCTCPQYKAGADCTEDVD-----ECAMANSNCEHAGKCVN--- 118
 QY 276 GDGS--GDCVGRGVGKCDMEITTDNCNG--TCHTSANCLDPDCKASKCAAGFRGNTV 332
 DB 119 TDGAFHECEKAGVAGPRCEMDI--NECHSDPCQMDATC-LDKIGFTCLCMPEK--GVH 173
 QY 333 C-TAINCET---SNGGCTKADCKRTTPGNRY-CVCKAGATGDTGYC-LEINPCLLENH 385
 DB 174 CELEVNCEQSNPCVNNQOCVDV-----NRFQCLCPREFTGP--VCQIDIDDC--SS 221
 QY 386 GCGDRNAECTQTGPNQAVNCLPRTYGDGKYSLINVCLTNNGGCSPEAFGNTYE---Q 441
 DB 222 TPCILNGAKCID-HPNGYECQCATGFTG-----ILCDEHIDCDPDP--CHHQCQCGDI 271
 QY 442 DQRTCTCKPRYTGDIYCRGSIYELPKNPSTSQYFFOLQHAHRELAGRPFTVYFAPLS 501
 DB 272 DSYTCINPGFM--GAITCSDOI--DECYSSPC-----LNDGRCIDLVN--GYOCNCQPGT 320
 QY 502 SSFHEPRIKMDQGLMSQVL-----RYHVGCCQLLLDNLKVTTSATTLQGPVSISV 556
 DB 321 SCLNCEINFDCCASNPCHGVGCVGGINRYSCV-----CSPGFTGCRNCIDI 366
 QY 557 SODTVFINNEAKVLSDIISTNGYIHVIDKLSPKNLLITPKDALGRVQLMTTVAANHG 616
 DB 367 DE---CASNPCKATGCIINDVNGFRICPEGRHPHPSQYNECLSNPCITGNCTGGLSG 423
 QY 617 YTKSKLIQDSGLSVTTDSITHPTVYVFWPTDKALEALPPQDQFLFNQDNKDKLKSTLK 676
 DB 424 Y---KCLCDAGWGVNCE-----VDKNECLSNPCQNGTGN---NLVNGY-- 462
 QY 677 FHVIRDSKALASDLPRASMKWTLQGSLSVRCGSDIGELFLNQMRFIHRLGLPEVG 736
 DB 463 -----KCTCKRGKFGVNCQV-----NIDECASNP--C--LNQICPEVDV 497
 QY 737 VAVGIDCLMNPITLGRGCDFTFTFDIPGEG-----SCITFP-----KCPL 777
 DB 498 SGYTGCHCL--PYTGKNCQIVLAPCSPNCPENAAVCKEAPNFESFCLCAAGWGQKRTV 555
 QY 778 K-----SKP---KGV-----KKCIYNPLPFRNRVCGCN- 804
 DB 556 DVDEICSPCMNNGVCHNTGGSYVCECPPEGSGMDCEEDINDCLANP-----CONG 606

QY 805 -LCTVVIQTPRC--CHGYEMPD-CQA-----CPGGDFPCCNNRGMKCRDLYTPMGQCLHTGF 857
 DB 607 GSCVDHWNFTSCQCHNPGFIDGKQOTDNNEC---LSEPCKNNGCTSD--YVNSYTCCTPAGF 662
 QY 858 NGTAGE-----LCWHGRFGP-----DCOPRSCSEHQ 884
 DB 663 HGVCENNNIDECTESSCFNGTCVDGINSFSLCPVGFTGFPFLHDINECSSNPCLNAGT 722
 QY 885 CDEITSGSECLTETGWTAAASCDPRTAVFVAVCPRA--CSVATCTENN---CYCNLWYEG 940
 DB 723 CVDEL--GYRICLPLGTTGKNQOT---LVNLCSNPPCKMKTGYEKARPHCLCPPEW-- 776
 QY 941 DGLTCTVDF--CK-----QNNGGCAKVAKCSQKGTQVSCSKKYGKGDYS 985
 DB 777 DGAVCDVLNNSKKAALQKQVPEHLQHSICINA-----GNTNHQCPGLGTYG--SYC 829
 QY 986 CIEIDPCADGVNGGCHENATCRMTGPKHKCEKSHYVGDSDPEPRLPLRDLQDNQ 1045
 DB 830 EEQIDDECA---SNPCOHGATCN--DFIGGYRCEVPGY--QGVNCEYE---VDEC--ONOP 878
 QY 1046 CHRPASCADLYFDQTYGVFHLRSPGLQYKLPFKAKKACKEAATLATYNSLYAQAK 1105
 DB 879 CONGCTCIDL-----VNHFKSCSPPTGRLCLCEINDECA-----CG 915
 QY 1106 YHLCASGWLSEGRVAYPTT---YASOKGANYVGIYDGSRANKSEMDVFCYRMK-DV 1160
 DB 916 PHCLNGCQCVNRIGGYTCRCLPFGAERCEBDINBCL---SNPQSE--GSLDCVQLKNY 971
 QY 1161 NCTCKAGYV-----DGFSC 1175
 DB 972 NCICRSAFYGRHCEFTLDVCPQPKPLNGLGTCVAVASNMPPDFTC 1014

RESULT 13

Notch homolog protein - sea squirt (Halocynthia roretzi)

C:Species: Halocynthia roretzi

C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000

C:Accession: T30201

R:Horii, S.; Saitoh, T.; Matsumoto, M.; Makabe, K.W.; Nishida, H.

Dev. Genes Evol. 207, 371-380, 1997

A:Title: Notch homologue from Halocynthia roretzi is preferentially expressed in the

A:Reference number: 220775

A:Accession: T30201

A:Status: preliminary; translated from GH/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-2352 <HOR>

A:Cross-references: EMBL:AB001327; NID:d1204472; PID:d1026501; PIDN:BAA25571.1

C:Genetics:

A:Gene: Notch

Query Match 6.0%; Score 474.5; DB 2; Length 2352;

Best Local Similarity 21.5%; Pred. No. 4.7e-21; Matches 275; Conservative 97; Mismatches 440; Indels 467; Gaps 67;

QY 135 NRCNNPTIIVRGEG-----KCSQAPCLELTKPLRETKCIYSTYFMKRSRYFIQCO 188
 DB 149 NECDTPDICONAGTCSNNDGYSCSVAGF-----EGNNECVNIDDCSGHS---CQ 196
 QY 189 --POCVRTIIRACMLASLANAKPAPGEVYKCALGT--ASVMDGV-----NGTGT 235
 DB 197 NGATCADAVSTYDCHC-----PAEWTGOYCTIIVDEGSLNNVAKRRDLQOTEGGFT 248
 QY 236 CQCGCLGNGTACE-----TCTEG--KYGIHCDAC--SC 265
 DB 249 CNCYVGGFTRDCCSENIDDCSNVACEFHNAKCIDQAGTEFCLCTPGNRILICHILDACISDPG 308
 QY 266 VHG--RCSGGLP--GDSGCDVGVGKVCMDMEITTDNC-----NOTCHTSANCLDP 314
 DB 309 ARGATCDTNPIITGHMCCPGMTDKDCSDI--DECSLGNPCEHNGCNNT----- 359
 QY 315 DGKASCKAAGFRGNGTVG--TAINACETSGGSCTKADCKRTTPGNRYCVCKAGYTGDI 373

```
Db DGESEICIVAGY--SGPRCEININECEPN--PCRNDATCLDMI--CNFNCVCMPTG--GI 412
360 VC-----LEINPCLENHGGCDRNAECQTGPNQAVNCLPRYTGDGKVCSL--INVCLTN 426
374 ICDEDIDECESPPC--ANGSTCLDEV-----NAVTCALGFTDD--CSONIDECAS 462
413 NGGCSPPFAFCVNTYEDODICTCKPDYTG-----DGIYCRGSIYGLPKNPSTSQYEFQ 479
427 PCNNKATC--IDKANAAYECBCAPGYTHVCHETNIDDCVINCHGSGCRDGVNTFYCCCL 519
463 LOEHAHVRLAGBPPTVAPLSSSPNHEPRIKMDQGLMSQVLYRHVVGCOQLLDNLK 539
480 LGYEGRK-----CQDITNECASSPCENGGTCDDELIGYTCG-----P 557
520 VTSATTLGGEPEVSISVSQDVFYINNEAKYSSDIISTNGVIHVIDKLSPNNLLITPKD 599
540 TGTSSSSE-----INPD 571
558 ALGRVQLNLTVAANHGYTKFSKLTQDSGLSVITDSITHPTVTPWPTDKALEALPPEQ 659
600 CVGNPCQYGTGV-----DGV-----DDYSCSCTP----- 595
572 DFLFNQDKKDKLSTLKHVLRDSCALASDLPRASMKTLQSGELSVRCGTG-----S 712
660 DIGEL-----FINEQCRFIHGLFEDVGAVGIDC-----LMMPTLGGRRD----- 755
713 DIOECSSNPCHEHYARRQHVHCIDAGY--OGENCETEINECASPPCHGACENKVAOPV 701
643 -FTTFEDIP-----GEGSCITFPKCPKSKRGVYK-----KCIYV 791
756 SHCDAGYGTACEIDINECATQPCQNGTCTGSIYNCACAPKATGVNCEITELSPCPN 761
702 PLPRFRNVEGQNLCTVYIOTPRCHGTFMPC-----QACRGGPTPCNNRGMCDLYTP 847
792 PC-----ENGATQESADYLAIVYCCQCEGFRGPLCATDINEC--VNSPCKNGGCGTNL--VP 814
762 MGOCILHTGFGNLTACELHWHGRFGPCQPRSCSEHGOCDEGTTGEGELCFETGWTAAASCD 907
848 GYQCTCSGSGFGTKCDDTI-----DDCSSNPLNGGQCLDLY--GSYKCLCLPFGFNNCQ 868
815 T-----PRAVPAVCT-----PACSVHATCTE-- 928
908 EEVNECAFPCCKNGGICTDYVNSVYCTCLSGFYSLDCEKNIEDSSSCMMNGTCVDGIN 928
869 NNTCVCNINTEGD-----GIT-----C 945
929 SYSCTANFTGDKCQANVNNCASLQCCNGGTCYVDSGDPKACVHGTYGTHCESLQMLC 988
929 TVVDFCKONNGCAKVAKCSQKGTQVSCSKKGYKD-----GYSCIEI--DP 991
946 TGPNTCK--NGS-----SCVTSNTVSCNCLGEGTIDCANPOVSCYTGASILGTAIVSDL 1041
989 CADGVNNGCHEHATCRMTGPKHKCECKSHYVGDVDEPBDPLDRCLQDNGOCHPDAS 1051
992 CLNG--GCHDITSTA-----HECSVAAGFTGSYCD--IDIDECA--SVPEKNAT 1085
1042 CADLYFOFTYGVFHLBPLQYKILTFPKAKEACKEAATATYVQSLYAKAKYHLCSA 1111
1052 CNDL-----INSYSICALGEGATCLTLDKDECASSP-----CKN 1120
1086 GMESEGRAVYPTTYASOKGAGNVGIVDYSRANKSEMDVFCYR-----MKDVNCTC 1164
1112 GGTCTDRL--NSFYCS--CLAGTEGVL--CEINDECEINCLNGVCLDGIIGFSQCC 1173
1121 KAGYVGGDFSCSGNLLQYL 1183
1165 PSQY--BGRRCQGDVNECL 1190
1174 RESULT 14
```

```
s42612
transmembrane protein precursor - zebra fish
C:Species: Brachydanio rerio (zebra fish)
C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 20-Sep-1999
C:Accession: S42612
R:Bierkamp, C.; Campos-Ortega, J.A.
Mech. Dev. 43, 87-100, 1993
A:Title: A zebrafish homologue of the Drosophila neurogenic gene Notch and its patter
A:Reference number: S42612; MUID:94128602
A:Accession: S42612
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-2437 <BIE>
A:Cross-references: EMBL:X69088; NID:9433866; PIDN:CAA48831.1; PID:9433867
C:Superfamily: unassigned ankryrin repeat proteins; ankryrin repeat homology; EGF homol
F:755-786/Domain: EGF homology <EGF1>
F:1023-1054/Domain: EGF homology <EGF>
F:1185-1216/Domain: ankryrin repeat homology <AN1>
F:1915-1947/Domain: ankryrin repeat homology <AN2>
F:1948-1980/Domain: ankryrin repeat homology <AN3>
F:1982-2014/Domain: ankryrin repeat homology <AN4>
F:2015-2047/Domain: ankryrin repeat homology <AN5>
F:2048-2080/Domain: ankryrin repeat homology <AN6>
```

```
Query Match 6.0%; Score 473.5; DB 2; Length 2437;
Best Local Similarity 20.3%; Pred. No. 5.6e-21;
Matches 303; Conservative 123; Mismatches 447; Indels 617; Gaps 84;

Db 135 NRCDNDITIVRGECC-----KCSQA--PCPLETKPLRETRKCIYSIY 176
146 NPCANG-----GQCSAFESHYICTPPNFHQTCRQDVNECAVSPSPCRMGNCINBEV- 198
177 FMGRSVETIGQPO-----CVRTITTRACWLASLAHNNKAPAEVKKMALGTASVWDGVN 231
199 -----GSYLCRCPPEYTGHCQR-----LYQCLPSP-----CRSGGTCVQTS-D 237
232 GTGTCCQGLGFGNFACE-----TCTE-----GKYGHCD-----QAC----- 263
238 TTHGTCSLPFTGQYCEHNVDDCTQHACENGCPCIDGINTYNCHDKDKMTQGYCTEDVDE 297
264 -----SCVHGRCGSGPLGDSGDCDGVGRGVKCDMEITTDNC-----NGTCH----- 305
298 CELSPNACONGTCHNTIGGFHCVCVNGMTGDDCSENI--DDCASAACSHGATCHDRVAS 355
306 -----TSANCLLDP--DKKASCKCAAGRGNGIVCT----- 334
336 FFCECPHGRTGLCHLDACISNPCQKGSNCDPTNPSGKAICTCPGY--TGSACNDID 413
335 -----AINAGETSNCGCSTKA--DCK-----RTTP-----GNR 360
414 ECSLGANFCEHGRCLNKTGSGFOCKLQGYEGPRCEMDVNECKSNPQONDATCLDQIGGF 473
361 VCVCAGAGTGGIYCLEIN-----PCL-----ENHG----- 387
474 HCICMPGY--EGVPC--QINSDDCASQPLNCKLIDKINSFCEGCPKGFSGSLCQVDVDEC 530
388 -----CDRAECTQGPONAVCNCLPKYTGDKRVSL--INVCLTNGGSPFAFCVPTDQD 442
531 ASTPCKNGAKCTD--GPNKYTCECTPGEFGIH--CELDINECAS-----SP--CHYGCR 579
443 QRI-----CTCRPDYTGDSIVCRGSIYGLPKNPSTSQYFQLOEHAHVRLAGPPTVFA 498
580 DGVASFVTCDCRPGYTG--LEETNI--NCLSLQPCNGSGTCDREKAYICTCPKG----- 630
499 PLSSSEHNEPRIKMDQGLMSQVLYRHVVGCOQLLDNLKVTTSATTLGGEPEVSISVSQ 558
631 --TTGVNCEINID-----CKRRCPCDYKCIDK----- 656
559 DTVFINNEAKYSSDIISTNGVIHVIDKLSPNNLLITPKALGRVQLNLTVAANHGYT 618
657 -----INGEYCEVCEPGYSGSMCNINIDCALNPHCHNGGTIDGV-----NSFTCLCPDGF- 706
```

QY 619 KFSKLIQDSGLLSTVTDTSIHPPVYVFWPTDKALEALPRPODFLPQNDKKDLKSTL---- 675
 Db 707 -----RATCLISQHNESSSNP-----CIIHSCLDQINSYRCVC 739
 QY 676 -----KFHVIDSKALASDLPRASAMKTLQSGEL-SVRCGTGSDIGELFLNE---QMC 725
 Db 740 EAGMNGRMCDININIECLSNPCYVNGSTCKDMTSGYLCTCAGSGSPRCOMNINIECASNP- 798
 QY 726 FIIHRLFDVGYAAYIDCLLMPRTLGGRCDFPTTEDIPEGC---GSC-----IFTPKC 775
 Db 799 -LNGSCSIDDVAGFCNCML--PYTGEVCENVLAPCSPRCKNGVCRESEDFQSGSCMC 855
 QY 776 PLKSPKPKGV---KCTINPLPRFRNVEGCONL-----CTVIYIOTPRCHGVF----- 820
 Db 856 PAGMGQGTCEVDINECVRNPC--TNGVCENLIRGGFQC-----RCNPGFALCEND 905
 QY 821 MPDQACP---GG-----PDPPCNRGMCRL---Y 845
 Db 906 IDDCFPNCSNCGVCODRYNGFVYCVCLAGFRGERCAEDIDECVASAPCRNGNCTDCVNSY 965
 QY 846 TFMGQCLHTHTNGTACEL-----CMHGRFGPD----- 873
 Db 966 T-----CSCPAGSSGINCEINTPDCTESSCFNNGTGVYGLISSVCVLPBFTGYCOHDVA 1021
 QY 874 -CQPRSCSEHGCDE-----GIMG-----SGRC 895
 Db 1022 ECDSPRCQNGGSGQDQVGYTKCTCPHGYTGLNCGSLVRMCDSPPCKNNGSGCMQOGASFIC 1081
 QY 886 LCETWTASCTPPPAVAVCT--PACSYHATCTENNTCV-----CNLYEG----- 940
 Db 1082 QCASGWTGTYCDVPVSYSCVAAARQGVSAVALCRHAGQCDVADGNTHLCRCQAGYTGYSYCO 1141
 QY 941 -----DGITCTYVDF-----CKONNGGCAVYAKSQKGTQV- 971
 Db 1142 EGYDECOQRPQNGATCT--DYLGYSCECVPGYHGMNSKELINELSL-QPCONGGTCTID 1198
 QY 972 -----SCSKRKYKGDGYSC-IEIDPCADGV-----NGG-CHEHATCRMTSPGRH 1014
 Db 1199 LVNTYKCKSPRGCTO--GVNCEIIDDCSPSVPLTGPGRPCFNQRCVDRV-----GGY 1249
 QY 1015 KCEGSHYVGDSDVDEPDLPLDRCLQDNQGCCHPDAS--CADLYFDITTVGVFHLRSLPG 1072
 Db 1250 GCVCPAGFVGE--RCEGD--VNECLSD--PCDPSGSGYCVOL-----INDFRCECRTG 1296
 QY 1073 ---GYKLFDKAEKACAEATITATYNQLSYAQKAKYHLCASGMLSESGVAAPPTYASO 1128
 Db 1297 YTKRCEIYFNCKDIDPKCKNGGTCAVASNTKHYICK---COPGYSGSS-----CEYDSS 1348
 QY 1129 KCGANVVGIVDGRSRANKSEMDVDFCYRMKDVNCTCKAGYVGDGFSC--SGNLQVLM 1184
 Db 1349 SCGS-----LRCRN-GATCVSHLSLPRCLCAFGFSHSEQOTRM 1385
 QY 1185 SFPSLTN-----FLTEVLAFSSKSARQO 1207
 Db 1386 DSPCLVNPICYNNGTCOPISDAFFYRCSCPANFGLCHILIDYSFSGQGR 1435

RESULT 15
 T13954
 MEGF6 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jul-2000
 C:Accession: T13954
 R:Nakayama, M.; Nakajima, D.; Nagase, T.; Nomura, N.; Seki, N.; Ohara, O.
 A>Title: Identification of high-molecular-weight proteins with multiple EGF-like motifs
 A:Reference number: Z14126; MUID:98360089
 A:Accession: T13954
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1574 <NAK>
 A:Cross-references: EMBL:AB011532; NID:g34449293; PIDN:BAA32462.1; PID:g3449294
 A:Experimental source: strain Sprague-Dawley; brain

Query Match	Best Local Similarity	22.4%	Prod. No. 3.5e-21	Matches 300	Conservative 99	Mismatches 460	Indels 478	Gaps
152	CSQAQPCPLETRKPLRETRKCIY	-----	SIYFMKRSYFI	-----	185			
Db	74	CAQAQAWC-----	IGQERITYYWYSRYQVATERTATYFRCCPGMSQKPGQEGCLSDVDEC	127				
QY	136	-----GCQPCVCFPIITRACKLASLAHNAKPA	---PGEYMKCA-LGTASWDG-----	230				
Db	128	ASANGCGEPCPCNTIVGGFYC-----	RCPPGYOLOGDKTCDVDYECRAHNNGCOHRGY	180				
QY	231	NOTCF--CQCGGLGF-----NGTAC---	ETCEBKGKGIHCDAQ--SCVHRGCS-----	271				
Db	181	MTPGSYLDECKPGRFLMTDGHRTCLAISSCYLGNGG--	COHQCVOLVYTHQRCORPOYOL	238				
QY	272	-----GQPLIGGS-----	CCDDVMGKVCMDMEI--TTNCCN--GTC	304				
Db	239	QEDGRKCYRSPSCAAGNGGCHTQELRLGLAHCCGHGYO--	LAADKRTCEDVDEBCLGLA	297				
QY	305	HTSANCLLDPDGKASCACAGFR--GNGTYC----	TAINACETSNCGCSTKADCKRTTP	357				
Db	298	QCAHCC-LNTGGSFPCVCHAGYELGADGRCYRIEMELIVSCEANGGCS--	HGCSHTST	354				
QY	338	GNRVCVCAGY--TGDIVCLEINPCLENHGGCDRAECTQTG--	PNOAVCNCLPKY-	410				
Db	355	G-PLCTCPRGYELDEDOATCIDLIDDC-----	ANSPPCCQACANTPGGIECSCFAGYR	405				
QY	411	-TGDCQVCSLLNVCILTNNGG-----	SPFAFCN--YTRODQICICXPDYAGDI	457				
Db	406	LNTDCCGEDVDEBCASGHGCEHHCSNLAGSFQFCAGTRLEDNRGCTSLSESYVD--	463					
QY	458	VCQSGIYELP-----	KNPSTSOYFPOLEH-AVRELAGPFFT-VFAPLSS	502				
Db	464	-----LDGRLEFPVRPLPIAVLRDELRLPDODVGAEBAAMAALREHNTLTKEFYCLDH	518					
QY	503	SNHEPRKLKMDQGLMSQVLRHVVGCOOLLNLNKVTTISATTLQGEPPSISVSODIYF	562					
Db	519	SGFHD-----	C-SLTCCDCR--MGCTCFPGDCDCDEPGMTGI	553				
QY	563	INNE-----	AKVLSDDIISTNGCIVHIVKLLS-----	PNLLIT-----	PRDALGRYL	605		
Db	554	ICNEFCPPDPTGKNCSSPCTCQNG--	GTCDDVLGACRCRPGVSAHCEDCPCRPFGYFKHC	611				
QY	606	ONULTTVAANHG--YTKFSLKLDQSGLLSVLTDSIHTPTVFWMPLDKALFALPREQDPLF	663					
Db	612	RR-KCHCANRGRCHRLYACCLCDPGLIGRFLHACPP-----	MAFG-----	PGSEBCLC	660			
QY	664	NQDN-----	KDKLSYLKFIHVRDSKALASDLPRASMT--LOGSLSVRCGTSGDIGE	716				
Db	661	EQSHRSCNPKD-----	GSCSKAGFGEGRCQAECSG-----	693				
QY	717	LFLANQMRFLHRIHLLPVGVA-----	YRIDCLNMP--TLGRCDDTLF	758				
Db	694	FFGFG-CR--HR-CTCPGVACDPVSGECRTQC	PGYQSEDCGQCEPVGTFVNCs--	745				
QY	759	TFDIDGEGCGSCITFP-----	KCLPSKPKRGYKKRCIYNPLRFRNVEGCONLCTVVI	810				
Db	746	GSC-SCVGA	PCHHVYTBECILCPGKGTBDDGACC-----	PEGMMIGLGCQEIC--	790			
QY	811	QTPRCCH-----	GYFMDQCQ--AC-----	827				
Db	791	--PACHEGASCNPETGTCLCPGFVSGHRCODTCSAGVYGTGQICRACANDGHCDDPTTGR	848					
QY	828	-----PG-----	GPTT--PCN--NRGMCRLTYTPMGQCLCHTFGNFACE	863				
Db	849	CSCAGGWTGSCQACRACDGSWHGPPCITHCCNSAGHGNCDAV--	SGLCLEAGYEGPRCE	905				
QY	864	-LCWHGRCPCPCQPCSCSEHG-QCDEGITGSGECLCTGWTAAACD--	TPTAFAAV--CTP	918				

```
Db 906 QSCROGYGSPSECEQKRCCEHSAACDH---VSGACTCPAGWRGSECEHACPAAGFFGLDCDS 962
QY 919 A--CSVHATC--TENNTCVCNLTNEGD-----GITCTVYDFCKQNNNGCAKVAKC 964
Db 963 ACNCSAGAPCDNAVYGSCLCPAGRWGPRCAQSCPPLTFGLNCSQICTC--FNGASCDSTG- 1020
QY 965 SQKGTQVSCSKKGYKGDYSCIEIDPCADGVNG--GCHHATCRMTG---PGKHKCECKS 1020
Db 1021 -----QCHCAPGWMGP--TCLQ--ACPPGLYKNCQHSCLCRNGGRCDPILGQCTCPE 1069
QY 1021 HYVGDGVCEPEBQLEPLDKCLQDNGQ---CHPDASCADLYFODTIVGVFHLRSPILGQTKL 1076
Db 1070 GWT--GLACENECILP-----GHYAAGCOLNCSLHGICDRLTG--HCLCPAG---W 1114
QY 1077 TEDKAKEACAKEAATITATYNNLSYAKAKYH-----LCSAGMLESG 1117
Db 1115 TGDRCQSSCVSGTFGVCHEHCACRKGASCHHVTGACFCPPGWRGPHCEQACPRGMFGBA 1174
QY 1118 --RVAYPTTYASQ-----KCGANVVGIV-----DYGSRAN-----KSEMDVFCY 1155
Db 1175 CAQRCLCPFTNASCHHVTGECRCPPGFTGLSCEQACQPGTFGKDCHEILCQCPGETW--ACD 1232
QY 1156 RMKDVNCTCKKAGYVGDG 1172
Db 1233 PASGV-CTCAAGYHGTG 1248
```

Search completed: June 17, 2002, 12:31:15
Job time: 281 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 17, 2002, 12:29:44 ; Search time 50.02 Seconds
(without alignments)
4949.135 Million cell updates/sec

Title: US-09-842-930A-2
Perfect score: 7861
Sequence: 1 SLPSLLRLRLQMPDYSIFRG.....WGHCQPMRSGQATTVYPR 1431

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5175.5	65.8	1192	4 Q9H7H7	Q9H7H7 homo sapien
2	4447.5	56.6	1069	4 Q9UF98	Q9UF98 homo sapien
3	3762.5	47.9	897	4 Q9NRY3	Q9NRY3 homo sapien
4	2820.5	35.9	2212	4 Q93072	Q93072 homo sapien
5	2810	35.7	2570	4 Q9NRT5	Q9NRT5 homo sapien
6	603.5	7.7	2189	5 Q9B105	Q9B105 elmeria ten
7	557	7.1	3857	11 Q88840	Q88840 mus muscula
8	555.5	7.1	937	5 Q9BLJ1	Q9BLJ1 clona intes
9	536	6.8	1664	5 Q9TVO2	Q9TVO2 caenorhabdi
10	519.5	6.6	2906	11 Q9WUH9	Q9WUH9 ratius norv
11	519	6.6	2653	5 Q25253	Q25253 lucilia cup
12	499.5	6.4	2872	11 Q9WUH8	Q9WUH8 ratius norv
13	489	6.2	2470	11 Q35516	Q35516 mus muscula
14	486.5	6.2	2524	5 Q9GPAS	Q9GPAS branchiosto
15	480.5	6.1	1111	5 Q9XWD6	Q9XWD6 caenorhabdi
16	475.5	6.0	1203	11 Q06008	Q06008 mus muscula

ALIGNMENTS

```

RESULT 1
Q9H7H7 PRELIMINARY; PRT; 1192 AA.
AC Q9H7H7;
DT 01-MAR-2001 (TREMREL. 16, Created)
DT 01-MAR-2001 (TREMREL. 16, Last sequence update)
DT 01-DEC-2001 (TREMREL. 19, Last annotation update)
DE FLJ00112 PROTEIN (FRAGMENT).
GN FLJ00112.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN;
RA Ohara O., Nagase T., Kikuno R., Okumura K.;
RT spleen.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK024503; BAB15793.1; -.
DR HSSP; P98066; ITRG.
DR InterPro; IPR000782; BIGH3_fasciclin.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002049; Laminin_EGF.
DR InterPro; IPR000538; Link.
DR Pfam; PF00008; EGF; 9.
DR Pfam; PF02469; Fasciclin; 2.
DR Pfam; PF00193; XLink; 1.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01265; LINKMODULE.
DR PRODOM; PD000918; Link; 1.
DR SMART; SM00181; EGF; 9.
DR SMART; SM00445; LINK; 1.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR NON_TER 1
FT SEQUENCE 1192 AA; 128738 MW; 5966BED4B83BD9C1 CRC64;
SO

```

```

17 474.5 6.0 2352 5 Q61240 061240 halocynthia
18 473 6.0 1574 11 Q88281 088281 ratius norv
19 469 6.0 2281 4 Q9UPL3 09UPL3 homo sapien
20 469 6.0 2321 4 Q916L8 0916L8 homo sapien
21 469 6.0 2321 4 Q9UM47 09UM47 homo sapien
22 465 5.9 2471 11 Q9QW30 09QW30 ratius sp.
23 461.5 5.9 1999 4 Q99400 099400 homo sapien
24 459.5 5.8 2003 4 Q00306 000306 homo sapien
25 453.5 5.8 3680 5 Q9VR08 09VR08 drosophila
26 452 5.7 2531 5 Q16004 016004 lytechinus
27 450.5 5.7 2471 4 Q04721 004721 homo sapien
28 450.5 5.7 2471 4 Q9H240 09H240 homo sapien
29 447 5.7 2319 11 Q91772 091772 ratius norv
30 443 5.6 2447 13 Q13149 013149 fuqu rubrip
31 437.5 5.6 1964 11 Q35442 035442 mus musculi
32 434 5.5 4135 6 Q18977 018977 bos taurus
33 431 5.5 3396 5 Q96KG6 096KG6 homo sapien
34 431 5.5 3396 5 Q9VM55 09VM55 drosophila
35 430 5.5 4288 4 Q9NPK9 09NPK9 homo sapien
36 424.5 5.4 1140 4 Q96KG7 096KG7 homo sapien
37 423.5 5.4 2104 5 Q964N4 0964N4 caenorhabdi
38 421 5.4 4006 11 Q35452 035452 mus musculi
39 420.5 5.3 2809 4 Q96JF8 096JF8 homo sapien
40 415 5.3 1242 13 Q90Y57 090Y57 brachydanto
41 414.5 5.3 1404 5 Q9VB65 09VB65 drosophila
42 413.5 5.3 2104 5 Q21281 021281 caenorhabdi
43 411.5 5.2 1722 5 Q19350 019350 caenorhabdi
44 410 5.2 1193 13 Q90819 090819 gallus gall
45 406 5.2 1214 13 Q90YD2 090YD2 xenopus lae

```



```

QY 556 VSODTFINNEAKVLSDDIISTNGIVHIDKLSKPNLLTPKDALGRVLOLTTVAANH 615
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 246 VSQSTVYINNKAKIISDDIISTNGIVHIIIDKLSKPNLLTPKDNKSGRILOLTTLATIN 305
QY 616 GYTFESKLIQDSGLSVTTSDTHFTVYFWPMDKALEALPPEQDFLNNQDKKLKSYL 675
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 306 GYIKFSNLIQDSGLSVTTSDTHFTVYFWPMDKALEALPPEQDFLNNQDKKLKLEYL 365
QY 676 KEHVIRDSKALASLPRASWMTLOGSELSVRCGSDIGELFLNEQCRFIHRGLLPVY 735
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 366 KEHVIRDSKALASLPRASWMTLOGSELSVRCGSDIGELFLNEQCRFIHRGLLPVY 425
QY 736 GVAVGIDCLLMNPVLGRCDFITFTFDIPQEGSSCIPTPKCPKSPKGVKKKCIYNPLPF 795
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 426 GVAVGIDCLLMNPVLGRCDFITFTFDIPQEGSSCIPTPKCPKSPKGVKKKCIYNPLPF 484
QY 796 RRVNBSGONLCTVYIQTTRCCCHGYFMPCQACPGGPDTPCANNRGKCRDLYTPMGQCLCHT 855
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 485 KRNEGGERKESLVYQIRRCCKGYFGRDQACPGGPDAPCANNRGVCLDQYSATGCKCKNT 544
QY 856 GFNGFACELCMHGRPRGPCOPRSCSEHGCDEGITSGBECLCEGTMTAASCDTPAVFAV 915
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 545 GFNGFACELCMHGRPRGPCOPRSCSEHGCDEGITSGBECLCEGTMTAASCDTPAVFAV 604
QY 916 CTAPACSVHATCTENNTCYCNINTEESDGIITCTVVDPCONNGCAKAVAKCSOKGTQVSCSC 975
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 605 CTAPACSVHATCTENNTCYCNINTEESDGIITCTVVDPCONNGCAKAVAKCSOKGTQVSCSC 664
QY 976 KKGKYGDSCTEIDPCADGVNNGGCHENATCRMTGPKHKCECKSHYVGDVDCPEBOLP 1035
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 665 QKGYKGDHSCSTEIDPCADGVNNGGCHENATCRMTGPKHKCECKSHYVGDVDCPEBOLP 724
QY 1036 LDRCLQDNGGCHADKACVLDLHPDPTTVGVFHLRSPLSGOYKLFEDKAKKEACKEATNTTY 1095
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 725 LDRCLQDNGGCHADKACVLDLHPDPTTVGVFHLRSPLSGOYKLFEDKAKKEACKEATNTTY 784
QY 1096 NOLSYAOKAKYHLCSAGMLESGRAVPTTYASQCKGAVNGIVDYGSRANKSEMDVFCY 1155
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 785 NOLSYAOKAKYHLCSAGMLESGRAVPTTYASQCKGAVNGIVDYGSRANKSEMDVFCY 844
QY 1156 RMKDVNCTCKKAGYVDDGSCSNLLOVLMSPFSLTNFLTEVLAFSKSSRGAFLKHLTD 1215
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 845 RMKG-----SAGLFQOLSSRPCIS-----RTPD 868
QY 1216 LSIKRTLEVPQNSGLPGKNSLGRDIEHHLTVNVVSFYNDLNGTEFLRTMLGSOLLTFPS 1275
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 869 LSIKRTLEVPQNSGLPGKNSLGRDIEHHLTVNVVSFYNDLNGTEFLRTMLGSOLLTFPS 928
QY 1276 QDQDLH-QETREVVDGRSILQMDIIAANGILHIISEPLRAPPTAATAHSGLTGIFCAVYL 1334
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 929 QDQDLH-QETREVVDGRSILQMDIIAANGILHIISEPLRAPPTAATAHSGLTGIFCAVYL 988
QY 1335 VTGAIALAAYSFRLKQRTTGFRFDOKRTLMWLLAS-----SSPRISQTLCKMPQRR 1388
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 989 VTGAIALAAYSFRLKQRTTGFRFDOKRTLMWLLAS-----SSPRISQTLCKMPQRR 1046
QY 1389 HPOSPPTPSP 1399
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1047 -PSYDPTDSE 1056

```

```

RESULT 3
Q9NRY3 PRELIMINARY: PRT: 897 AA.
AC Q9NRY3.
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CD44-Like PRECURSOR FELL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_taxid=9606;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RA Tao O., Zhang W., Cao X.;
RT "Molecular cloning and characterization of human FELL sharing homology
    with CD44.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF160476; AAF82398.1; -.
DR HSSP: P98066; TRSG.
DR InterPro: IPR000782; BIGH3_fasciclin.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000538; Link.
DR Pfam: PF02469; Fasciclin; 1.
DR Pfam: PF00193; Xlink; 1.
DR PRINTS: PR01265; LINKMODULE.
DR Prodom: PD000918; Link; 1.
DR SMART: SM00181; EGF; 5.
DR SMART: SM00445; Link; 1.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 4.
DR PROSITE: PS01248; LAMININ_TYPE_EGF; 1.
KM EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.
SQ SEQUENCE 897 AA; 97385 MW; E8920AF36101E388 CRC64;

```

Query Match 47.9%; Score 3762.5; DB 4; Length 897;

Best Local Similarity 76.1%; Pred. No. 0; Mismatches 114; Indels 11; Gaps 4;

Matches 676; Conservative 87; Indels 11; Gaps 4;

```

QY 519 MSQVLRHYHVGCOQDLNLLDKVYTSATPLQGEPRVSIYSQDTPFINNEAKVLSDDIISTN 578
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 1 MPQVLRHYHVGCOQDLNLLDKVYTSATPLQGEPRVSIYSQDTPFINNEAKVLSDDIISTN 60
QY 579 GYIHYIDKLSKPNLLTPKDALGRVLOLTTVAANHGYTKRSKLIQDSGLSVTTSDTH 638
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 61 GYIHYIDKLSKPNLLTPKDNKSGRILOLTTLATINNGIKFSNLIQDSGLSVTTSDTH 120
QY 639 TPVTFWMPMDKALEALPPEQDFLNNQDKKLKSYLKEHVIRDSKALASLPRASWMT 698
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 121 TPVTFWMPMDKALEALPPEQDFLNNQDKKLKLEYLKEHVIRDSKALASLPRASWMT 180
QY 699 LQGSLSVRCGSDIGELFLNEQCRFIHRGLLPDVGYAVGIDCLLMNPVLGRCDFIT 758
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 181 LQGSLSVRCGSDIGELFLNEQCRFIHRGLLPDVGYAVGIDCLLMNPVLGRCDFIT 240
QY 759 TPFDIPEGSSCIFTTPKCPKSPKGVKKKCIYNPLPRFRNBSGONLCTVYIQTTRCCCHG 818
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 241 TPFDASGEGSCVNTPSCPRMSKPKGVKKKCIYNPLPRFRNBSGONLCTVYIQTTRCCCHG 299
QY 819 YEMPCQACPGGPDTPCANNRGKCRDLYTPMGQCLCHTFGNGTACELCMHGRPRGPCOPRS 878
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 300 YEMPCQACPGGPDTPCANNRGKCRDLYTPMGQCLCHTFGNGTACELCMHGRPRGPCOPRS 359
QY 879 CSEHGCDEGITSGBECLCEGTMTAASCDTPAVFAVCTPACSVHATCTENNTCYCNINLY 938
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 360 CSEHGCDEGITSGBECLCEGTMTAASCDTPAVFAVCTPACSVHATCTENNTCYCNINLY 419
QY 939 EGDGITYVVDPCONNGCAKAVAKCSOKGTQVSCSKKGYGDSCTEIDPCADGVNG 998
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 420 EGDGITYVVDPCONNGCAKAVAKCSOKGTQVSCSKKGYGDSCTEIDPCADGVNG 479
QY 999 GCEHATCRMTGPKHKCECKSHYVGDVDCPEBOLPDLRCLQDNGGCHADKACVLDLHPQ 1058
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 480 GCEHATCRMTGPKHKCECKSHYVGDVDCPEBOLPDLRCLQDNGGCHADKACVLDLHPQ 539
QY 1059 DTPVGVFHLRSPLSGOYKLFEDKAKKEACKEATNTTYNOLSYAOKAKYHLCSAGMLESGR 1118
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 540 DTPVGVFHLRSPLSGOYKLFEDKAKKEACKEATNTTYNOLSYAOKAKYHLCSAGMLESGR 599
QY 1119 VAYPTTYASQCKGAVNGIVDYGSRANKSEMDVFCYRRKDVNCTCKKAGYVDDGSCSN 1178
    |||:||||:||||:||||:||||:||||:||||:||||:||||:||||:||||:
Db 600 VAYPTTYASQCKGAVNGIVDYGSRANKSEMDVFCYRRKDVNCTCKKAGYVDDGSCSN 659

```

OY	1179	LLOVAMSPSTLNFLETVLAIFSKSARCAQFLKHLTDLSIRGTLFVPWONGSLPGKKSLSG	1238
Dd	660	LLOVAMSPSTLNFLETVLAIFSSNSRGARFALEHLLDLSIRGTLFVPWONGSLGENETLSG	719
OY	1239	RDIHEHLNVNVSFPNDLVNGTFELTLMGLSQDLLITFSODLI--OETRFDGRSLIQLMDII	1297
Dd	720	RDIHEHLNVSMFEFNYNGTTLTQTGVOSKRLILIASODPLQPTETRFVDGARILLQWDIEF	779
OY	1298	AANGLIHIISPELRAPPTRATTAHSGLGTGICAVLYVTGAIALAAYSIFBLKORTTGFO	1357
Dd	780	ASNGIHIVISRKLAPPAVPYLTHHGGLAGIFFAILLVGAAVALAAYSFYRINRRITGFQ	839
OY	1358	RPDOKRTMLMSWL-----SSPRTISQTCMRPQRHRHPQSPVPVPSQ	1399
Dd	840	HFESEEDINVVALGKOQPENISNPLEYESTTSAPPE---PSYDPFTDSE	884
RESULT	4		
ID	093072	PRELIMINARY;	PRT; 2212 AA.
AC	093072		
DT	01-FEB-1997	(TREMBLrel. 02, Created)	
DT	01-FEB-1997	(TREMBLrel. 02, Last sequence update)	
DT	01-DEC-2001	(TREMBLrel. 19, Last annotation update)	
DE	KIAA0246	PROTEIN (FRAGMENT).	
GN	KIAA0246.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX	NCHI_Taxid=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BONE MARROW;		
RX	MEDLINE=97191544; PubMed=9039502;		
RA	Nadase T., Seki N., Ishikawa K., Ohira M., Kawarabayasi Y., Ohara O., Tanaka A., Kotani H., Miyajima N., Nomura N.;		
RT	"Prediction of the coding sequences of unidentified human genes. VI. the coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by RT analysis of cDNA clones from cell line KG-1 and brain."		
RL	DNA Res. 3:321-329(1996).		
DR	EMBL; D87433; BAA13377.1; -		
DR	HSSP; P98066; ITSG.		
DR	InterPro: IPR000782; BiG3_fasciclin.		
DR	InterPro: IPR001128; Cyt_P450.		
DR	InterPro: IPR000561; EGF-like.		
DR	InterPro: IPR002049; Laminin_EGF.		
DR	pfam: PF00008; EGF_13.		
DR	pfam: PF02469; Fasciclin; 3.		
DR	PRINTS: PR01265; LINKMODULE.		
DR	ProDom: PD000918; Link; 1.		
DR	SMART; SMO0180; EGF_Lam; 1.		
DR	SMART; SMO0445; Link; 1.		
DR	PROSITE: PS00086; CYTOCHROME_P450; UNKNOWN_2.		
DR	PROSITE: PS00022; EGF_1; UNKNOWN_5.		
DR	PROSITE: PS01186; EGF_2; 13.		
DR	PROSITE: PS01248; LAMININ_TYPE_EGF; 2.		
KM	EGF-like domain; Glycoprotein; Laminin EGF-like domain; Repeat.		
FT	NON TER		
FT	1		
FT	2212		
SQ	SEQUENCE	2212 AA; 237451 MW; 4A95460504129134 CRC64;	
Query Match	35.9%; Score 2820.5; DB 4; Length 2212;		
Best Local Similarity	40.2%; Pred. No. 1e-229;		
Matches	563; Conservative 219; Mismatches 554; Indels 65; Gaps 20;		
OY	5	LITRELPMDVISIFRGIYHYNLASIESADAYVFVNNEAIENTIREKKATSLAKEDIL	64
Dd	774	LLOGLDLVPALFISPELLQHGLVPOIEAAATYTTFVPTNRILE--AGNGSHLDADV	830

QY	65	RHYVVLGEKLLKNDLHNGMREHMLGFSYLLAFLLRNDOLYNEAPINTNTVATKGVIH	124
Db	831	RHHVVLGELALSMETLLRRGGHRRNSTLLGPAHVIYVNHSGPEVNHVPLEBGMLEADGRSLI	890
QY	125	GLEKVLLETQKRNCRNNNTIIVRGEGCGSSQOAPC-----PLETKPLMETKCTIYSIFPMGK	180
Db	891	GLSGVLTVYGSSRCLHSHAEALRECVNCTHFRCTQGOLODTP---RKSVCYBSGFBS	947
QY	181	RSVFICOPQCVRTIITRACVLASLHNAKPARAEVKMCALGTASVMDGVNGSTGTCQGL	240
Db	948	R-----GGSYCTAKKIQVPDDCPGFEGFTLCRPPCGGGAGVCSGGQOCODRFLGSGECHQHE	10030
QY	241	GENGTACTCEGKYUGJHCDQASCVCYGRSOGPLDGGSDDDVGMRGVCMETTTTNC	300
Db	1004	GPHGTACEVCLGRKGRNCTGVCDCAHGLCDBELOGDSCVCYVGMQGLRCDQKITTSPQC	10633
QY	301	NGTCHTSANCLLDDPGKASCCAAEFGKNGTVCJAINAGETSNGGCSTKADCKRTTPGNR	360
Db	1064	PRKCPNPANCYQDSAGASTACAGAGYSGNIGFCESEVDPAHNGCGSPHANTKYAPQGR	11232
QY	361	WCVCAGTGTGGIYCLINCLNLEHNGGCDNACSTQGNQAVNCNLPKYTSDG-KVSL	419
Db	1124	TCTQDQGMGGGELCOEINSLHHGGCHHAECITPGQOVSCSREGSDGIRTEL	11833
QY	420	INVLCTNNGGSPFAFCNYYEDORLCTCKPDT-GDGJVCRSYIGELPKPMSYQYF	478
Db	1184	LDPCSKNNGGSPVATCKSTGDDGRTCTDPAHTVDDGLTCRARVGLERDKNAS--FF	12414
QY	479	OLQEHAVNELBPPFTYFAP-----LSSFFNH-----PRKHQDQGLMSOYLRYHVVGCQOL	533
Db	1242	SLRLLEYKELKGDGPPITPVHADLMNLSQDELARIHAHO-----LVFRHVVGCRL	12966
QY	534	LDLNLKVTTSATTLQGEFVSIYSQDPTFENNEAKVLLSDIISTNGIHVHDKLSPKNL	593
Db	1297	RSEDLLEGVATFALSGHPLRSEREGSYIYNDFARVSSDHEANGVILHFDIVLLPREA	13566
QY	594	LITPKDALGRVLONLTVVAANHGYTKFSKLIQDSLLSYTDSIHTPVTVFPTDKALEA	653
Db	1357	LHMERDDAPPIRRRVNTAAAGOFKJFISGLKVAAGLLPLREASHRPFMLMPTDAARA	1416
QY	654	LPPEODPLFPMODMKDKLSYLFKFNHVRDASKALASDLPSPASAKTLOGSELVSRGTGSD	713
Db	1417	LPDRQAMLYHEDHRDKLAALRGHMRNVEALASDLPNLGRPLRTMHGRIPIFSC-SMTR	1475
QY	714	IGELTFEOMCRFTHRGSLFDVGVAVGIDCLLNPITLGGRCDTFTTFDIP-GEGSGCIFT	772
Db	1476	PGELMWGSEDDARIYQRIHPEGGIAGYIDDLPEPGLARGCHETFTRPLRYLWTCISGLE	15353
QY	773	PKCPLKSPKVKKKK-----IYNLPFR-----RWV-----EGCQNLCTV	808
Db	1536	PRCPGSEQSOBSPACMRFPKETSPLHSLGRSVNHPSLMGRPOGLGMRGRNCVT	15955
QY	809	VLOTPRCCHGTFMEQCOACGAGGPPPTPCNNRGMCRDLYTTPMGQCLCTHGFNGTACELCMHG	868
Db	1596	TTRMKDSCCPGHYGSEQCAQCPGGSPCSDRGVCMQDMSGGOCLCHSGPAGTACELCARG	16555
QY	869	RFGDQCPRSCSEHNGQDEBITGSGBELCTGTGAASCDTPPAVFAVCTPRACSVNATCTE	928
Db	1656	AFGPRCQACRCTVHGRKDEGLAGSGSCFCDEBGTGRCVEOULELPVCHPRPARAVAGRA	1715
QY	929	NNTCVCNLYNGDDITCTCVDPFCQONNKGCAKAAKOSKOTGVSCSKGKYGGDSCISLE	988
Db	1716	GNSCECSLGTGSDRGVCTVADLCODHGGGCSSEHANCISOYGTWYTCCLPDIYGGDMSGRA	1775
QY	989	IDPCADGVNGGCHBHATCRMTGPKHAKCECKSHYVGDGVDC--EPDQLIDRCLQDNQCH	1047
Db	1776	RNPCTCHRGSGSESHANCLSTGLLTRRCCEHAGYVGDGIQCLESEPPRYDRLQAPRPC	18353
QY	1048	PDASADLYPDQDTVGVFHLRSPYLGQKLLTFDKAKAEAKAATATYTNQLSYQAQKXVH	1107
Db	1836	SDAMCTDTHFQEKRAGVFHLOATISGPRGLNFSAEAAACEAGVILASFPOLSAAOOLGFH	1895

DB 2253 LCLMGLANGSTAHVPVADCCGNGRVGVSLGARKMLSERMDAYCFRVDVACRCHNG 2312
1168 YVGGGFS-CGSQLQVLMFSLTFLEVLAFSSSARGAQLKHLFDLSIRGLFVPO 1226
DB 2313 FVGGISICNCKRLDLVLAATANFSTFYGMLGYANATRGIDFDELDELTYKTLFVAV 2372
QY 1227 NSGLPGNLSGRDIEHLFTVNVNSFYNDIVNGTFLRTMLGSOLLTFESODLHOETREV 1286
DB 2373 NEGFDVNNLTGSPDELHASNATLISAN-ASGKLLPAHSGSLTILSDAPDNSSMAVVA 2431
QY 1287 DG---NSILQMDITANGILHIITSEPLRAPP-----TAATRAHSGDCTGTFCAVYLVG 1337
DB 2432 PGTVVVSRITVMDIMAFNGIITHALASPLIAPPOQAVLAPPAVAVAGV-----AVLAAG 2487
QY 1338 AL-ALAAVSYERLKKORTTGF 1356
DB 2488 ALLGLVAGALLYLRARCKPTGF 2508

RESULT 6

09B105 PRELIMINARY: PRT: 2189 AA.

AC 09B105;
DT 01-JUN-2001 (TREMBlrel, 17, Created)
DT 01-JUN-2001 (TREMBlrel, 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel, 19, Last annotation update)
DE MICROME PROTEIN 4.
CN Mic4.
OS Eimeria tenella.
OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;
OC Eimeria.
OX NCBI_TaxID=5802;
RN [1]
RC SEQUENCE FROM N.A.
RC STRAIN=HOUGHTON;
RA Tomley F.M., Billington K.J., Bumstead J.M., Clark J.D., Monaghan P.,
RT "Emi4: A microneme protein from Eimeria tenella that contains tandem
RT arrays of epidermal growth factor-like and thrombospondin type-1
RT repeats.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ306453; CAC34726.1; -.
DR HSSP: P33555; IEMN.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00008; EGF_24.
DR SMART: SM00181; EGF_30.
DR SMART: SM00179; EGF_CA_30.
DR PROSITE: PS00010; ASX_HYDROXYL, 22.
DR PROSITE: PS01186; EGF_2; 18.
DR PROSITE: PS01187; EGF_CA; 18.
DR PROSITE: PS50092; TSP1; 1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 2189 AA; 224388 MW; C272A420B94FCB2D CRC64;

Query Match 7.7%; Score 603.5; DB 5; Length 2189;
Best Local Similarity 24.3%; Pred. No. 2.3e-41;
Matches 296; Conservative 114; Mismatches 459; Indels 347; Gaps 73;

QY 132 IOKNCDNNMTTIVRGECKGCSQAP-----C-----PLETRKPLRE--TRKCIYSIFM 178
DB 464 MEGNACSDID-----ECSEASTEIPENCNCYNTGESSFLEKPGVELVDGKCV-KIDF- 515
QY 179 GKRVSFICQPCQVRTITTRACMIASLAHNAKPAPGEVKMCAIGTASVWDGVNGTGTCCQ 238
DB 516 -----CAR-----GAC-NSLAHCKENPEGTALIC-----TCIA 542
QY 239 GLGFGNCTACETGEGKYGTHCDQACSCV-----HRCSCGSPGSGSCDCCVGNWR-- 287
DB 543 GYSGDGT-----QG-----HCDDIDECIAENDCTPADGGIGENTVGSYTKCAAGYQOD 593

QY 288 GVKCDMEITITDNC-NGT--CHTSANCLIDPDGKASCCKACAGFRNGIVCTA1NACETSNG 344
DB 594 GNSC---TIDECANGCHINCHASATC-TNFGSFEACACNAGSGVGCNDVDECTAD 649
QY 345 GCSRADCKRTTGCNRVVCVCAKY-TGDGIVCLEINPCLENHGGCDRAECTGTGPNQAV 403
DB 650 DCGENTLTCNNTV-GSEFECTCMAFEADAATKCTIDECASGTHCTSTACTNTNA-GSFT 707
QY 404 CNGLPKRYTGDCKVCSLIWVCLTNNGGSPAPFCNTEEDORI-CTCKPDYGDGIVCRGS 462
DB 708 CECNPFSGDGHKCEDVFPCCGGLHDCVNAHBCSSDNTTFTKCTGIGYEGE----- 761
QY 463 IYGLPKNPSTSOYFFLOEHAVERLACGPPTVAPLSSFFNHEPRIKMDQGLMSQV 522
DB 762 HGE-----NCGQDIDECADQAI-----CGENTVCTNTPGSFE-----CACV 797
QY 523 LRIYHVGCOQLLDNLKATTSATIL-----QGEVVSISVQDIYFINNEA----- 567
DB 798 EGFAVAVGAK-----LKGATSLTICIDDECNDASKNTCATSADGSCKNTAGSECSCLP 851
QY 568 -----KVLSDITSTNGVI--HVIDKLSPKNLLIPKDALGVLQNLTTVAANHGYT 618
DB 852 GPGDGHSCCTIDECATGCGCEHA-----TCENTAGS--YNTCEA---GYT 894
QY 619 KFSKLIQDSGLSVITDSIHPTVFWPTDK-----ALEALPPOQDFLENDQNKD 669
DB 895 -----QODGAVGCCIDIDECASATVLPANATCVNTEGSEYFECVP---GYRHTENCT 944
QY 670 KIKSYLKFHVLRDSKALASDLPRASAMKTLOGSELVRCGQSD-TGELFLNEOMCRFIH 728
DB 945 K-----IDFCSERGCNANASKEND-----GTEAICHSQYEGNCE---GEECKKID 991
QY 729 RGLLFDVGVANGIDCLMNPILGCRCDTFTTEDIGEGSCCIFPKCPDLKSPK-GYKK 787
DB 992 -----EC-----SYGEPCKDF-----GEGGVYDVSFGSCSCATGFIKR 1027
QY 788 CIYNLPPRRNVGCONLCTVVIOTPRCHGFMPCOACPGPDTPCNNRGMCHDLVTP 847
DB 1028 CT-----CQDI-----DECLDCK-MNTC--APVG-----GICNTVGS 1057
QY 848 MQGCLCHTGF-NGTACELCHMGRGPRCQPRSCSEHQCDEGIT-----GSGECLCTG 900
DB 1058 F-TSCCAAGFTGDLTCE-----DIDECATAH-TCDPNATCVNTVSGSECGCKEG 1106
QY 901 WTAACDPTFAFAVCTP---ACSVHATCTENN---TCVCNLANE--GDGTTCTVDFEC 951
DB 1107 FSGDG-HICTDIDECADNLNKNCDHKGICQNGTGSYTCGCRPGVSLAADFTCDNDEC 1165
QY 952 KONNGCAKVAKCSOKGTQVSCSCCKKGYKGDYSCTEIDPCADGVNGGCHENATCRMTGP 1011
DB 1166 AAGTATGERSFCVDPJQYSKCECKNGYRQCGEDVDVDECEADVH-TCSEHATCTNT-E 1223
QY 1012 GKHKCECKSHYVGDVDDEPEQLPDRCLQDMGCHPAPASCDLTFQDTTVVHFLRSLP 1071
DB 1224 GSHTCNEGYGGDKCKEKTIVP--C--DNSPGNNAMC-----EATLADSYCTCKA 1272
QY 1072 GOYKLTIPKAKPACAK-EAATATATYNOLSYAOKAKYH-----LCSAGMLESGRAYPTTY 1125
DB 1273 G-----YEMKDACVYDIDECOSGTHNCBPHADCSNTDGSFTCTGSGGTVGVTLCEDV-- 1325
QY 1126 ASQKGANVVGIVDYSRANKSEMDVFCYRMKDYN-----CTCAGYVGDGFS 1174
DB 1326 --DECAGNHAG-----C-----D1NACTNVPSGFTCECKSGSGFGDDE 1362
QY 1175 CSGN-LLOYLMSFBLTNLEVLAFSSSSARG-----QAFKLKLTLSIRGLTFVPO 1227
DB 1363 CTEKVLILPGQIHCDSDWTAW-TECTAETKOSTRKCVALLKVEVKLCPADISAGCELGEW 1421
QY 1228 SGLPG-NKSLSGDIE 1242
DB 1422 SSCPGVNNLISHRAE 1437

```
RESULT 7
088840 ID 088840 PRELIMINARY; PRT: 3857 AA.
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE MUTANT FIBRILLIN-1.
GN FBN1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B10.D2;
RX MEDLINE=98069008; PubMed=9405934;
RA Bona C.A., Murai C., Casares S., Kasturi K., Nishimura H., Honjo T.,
RA Matsuda F.;
RT "Structure of the mutant fibrillin-1 gene in the tight skin (TSK)
RT mouse ";
RL DNA Res. 4:267-271(1997).
DR EMBL; AF007248; AAC62317.1; -.
DR HSSP; P35555; IAPJ.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR002557; Chitin_binding.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000822; ZnF-C2H2.
DR Pfam; PF00008; EGF; 64.
DR Pfam; PF00683; TB; 12.
DR SMART; SM00494; ChrbD2; 2.
DR SMART; SM00179; EGF_CA; 60.
DR SMART; SM00001; EGF_Like; 4.
DR PROSITE; PS00010; ASX_HYDROXYL; 61.
DR PROSITE; PS00022; EGF_1; UNKNOWN_2.
DR PROSITE; PS01186; EGF_2; 50.
DR PROSITE; PS01187; EGF_CA; 61.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 3857 AA; 418301 MW; 5BC0618BC527E04C CRC64;
SQ

Query Match 7.1%; Score 557; DB 11; Length 3857;
Best Local Similarity 20.6%; Pred. No. 5e-37;
Matches 296; Conservative 136; Mismatches 403; Indels 600; Gaps 79;

QY 132 IOKNCDNNNDTIIVRGEC---GKCSQAPCPLETKPLRETRKCI---YSIYFMG---- 179
DB 1156 IDINCELSANLCPRHGKRVNLIGK--YOCACNPGYHPHTRDLFCVDIDECSTIMNGCETFF 1213
QY 180 ----KRSVFICGOCPCVPTTITRACWLASLAHNAKPADEYVMCALGTASVWDG-----V 230
DB 1214 CTNSDGSYECSCOPFALMPDRSC-----TDIDECF-DNPNICDGCQCTNI 1259
QY 221 NGTGCOCGLGFGNTA-CETCEGKYGGIHCD-QACSVHGRSCQPLRGSGDDCVGMWG 288
DB 1260 PGEYKCLCYDGFMASEDMKTCVDVN--ECDLNPNICLSGTC-ENTKSGFTCHDMGYSG 1315
QY 289 VK----C---DMEITTTDNC--NGTCHTSANCLDIPDGKASCKCAAGFGNGTCTATNA 338
DB 1316 KKGKGTCTDINECELGAINCGHAYCTMTA-----GSFKSCSPGWTGDDIKCTDDE 1368
QY 339 CETSNGGCGSTADCKRTTPGNRVCAKAGYTGDIIVCLEINPCLNHHGGCDRNAECTQTG 398
DB 1369 CSNGTHHMSOHADCKNTM-GSYRCLCKDGYTGDFCTDIDECSENILNLCG-NGOCL-NA 1425
QY 399 PQAAYCNC-----LPKVTDDGV-----CSLINVCL-----TN 426
DB 1426 PGGYKCECDMGFVP--SADGKACEDIDECSLPNICVFGTCINLPGILPCECEIGYELDRS 1483
```

```
QY 427 NGGCSPEAFEC-----NYTEDORICTCKPDY----TGDGIV--CRGSIY 464
DB 1484 GGNCTDVNECLDPTTCISGNCVNTPGSYT-----CDCPDEELNPTRVGCVDTCSGNCY 1537
QY 465 -----FQQLQEHAVRELAGP-----GELPKNPSTSOY----- 476
DB 1538 LDIPRPGNDGYACSNELGVGVSKASCCSLGKAWGPCELCPSVNTSEYKILCPGEGGF 1597
QY 477 -----FQQLQEHAVRELAGP-----GPTTVAPISSSNHPRKID----WD 514
DB 1598 RNPPTVILIEDIDE--COELPGLCOGKCINTFGSPQCHCPGYIYNPTRVCDVNECE 1655
QY 515 OQGLMSQVLRHYVHVCQOQLL-----DNLKVTTSATLQGEPPYSISVQPTVFEN 564
DB 1656 TPGLCGPGTCVNTGVNATYCTICPPDYMQVGNNAEYALCSSG--PGMTSAGTDI---- 1710
QY 565 NEAKVLSSDIISTNGYIHWIDKLSPKNLLTPK-----DALGRVLQNLTYVAANH 615
DB 1711 NBC-ALDPP-ICPNGIC-----ENLRGYKCIQNSGYEVDITGKNCVDINECVLN- 1758
QY 616 GYTKFSKLIQDSGLSVITDSTHTPTV-----VFMPTRKALEALPPRQODFLFND 666
DB 1759 -----SLCDNG-----QCRNTPGSEYVCTCPKGFVYAPDLKTECIDECSSPCTNCGV 1806
QY 667 NRDKLSYLV----KPHVIRDSKALASDLPRASAMKTLQSELSVRCGTSIDIGELFLNQ 722
DB 1807 CANSQSFICECSPESTLDPTTICETITKGCWQIV-----IDGREI--NINGATLKSE 1860
QY 723 KCRFTHRGLLPVGVAYGIDCL--MNPTLG-----GRCDFTFTTPTDIPGEC 766
DB 1861 CCS-----SLGAAMGSPCTICQDLPICGKGFIRKIQCEDINECEVF-----PGVC 1907
QY 767 -GSCLEF----- 772
DB 1908 KGLCLVNSKGSFKCECPNMGMTLDATGRLCLDIRLETCFLKYDEDECTLPDIAGRHRMDACC 1967
QY 773 -----PKPLKS-----KPKGVKKCIYNPLPRRNVNGCO--NLCT- 807
DB 1968 CSVGAAMGTBEECEBCEPLRNSREYBELCPRGPFATKRDITNGAPRFEDIDECKMIPSLCH 2027
QY 808 -----VVIQTP-----RCHGAY 819
DB 2028 GACRMTIGSFKRCRDSGFALDSEERNCTDIDECRIISPDLGKGGCVNPGDECECDEG 2087
QY 820 ---FM-----PPCQ---ACPGGPRPPCNNR-----GMCRLD-- 844
DB 2088 ESEFMKMKNCMDIDECQRPDLICRG--ICNNTBGSYRCECPRGHQLSPNISACIDINE 2144
QY 845 -----YTPMG-----QCLCHTGFGNGT-----ACE-ICWH--GR 869
DB 2145 CELSANLCPHGKCVNLICKYOCACNPGYHPHTRDLFCVVIDECSTIMNGCETFFCTNSDS 2204
QY 870 FGPPDCOP-----RSCSEHGQDE-----GITSGECLCTGTAA----- 904
DB 2205 YECSCQPGFALMPDRSCDIDECEDNPNICDGGCTNIPGEYRCLCYGFMASEDMKTC 2264
QY 905 -----SCQPTTAVFAV-----CTPA-----CSVHAT 925
DB 2265 VUVNECDLNPNICLSGTCENTGTGSCFICHCMDGYSGKKGKTCGTCTDINECTIGAHNCDBRAV 2324
QY 926 CTENNT-----CVCMNLNVEGDIITCTVVDFCKONNGGCAKVAKCSQKGYOVSCSKKGYK 980
DB 2325 CT--NTAGSFKSCSPGWIGDGIKCTDIDECSSGTHMGQHDCKNTKMSYGLCKDGYT 2382
QY 981 GDGYSCIEIDPCADGV-----NGCCHNATCRMTGPKHKECKESHV--GDGVDCPEPOL 1034
DB 2383 GCGFTCTDIDECSENLNLCNGQC-----LNAGGAYCECDMGFVPSADKAGE----- 2431
QY 1035 PLDRCLQDN-----GQCHPDASCADLYFDPTVGVFHLRLSPLOQYKLTTPKAKAEC 1085
DB 2432 DIDECSLPNICVFGTCH-----NLPGILPCECEIG-YEL--DRSGNC 2471
```

```

RESULT 8
ID 09BLJ1 PRELIMINARY: PRT: 937 AA.
AC 09BLJ1:
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE CI-METAL.
GN CI-METAL.
OS Clona intestinalis.
OC Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Phlebobranchia;
OC Clonidae; Clona.
OX NCBI_TaxID=7719;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21347414; PubMed=11455433;
RA Nakayama A., Satou Y., Satou N.;
RT "Isolation and characterization of genes that are expressed during
  Clona intestinalis metamorphosis.";
RL Dev. Genes Evol. 211:184-189(2001).
DR EMBL: AB041857; BAB40596.1; -.
DR HSSP: P00742; IHCG.
DR InterPro: IPR00152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF-2.
DR InterPro: IPR001881; EGF-Ca.
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR001491; Thrombomodulin.
DR Pfam: PF00008; EGF_5.
DR PRINTS: PRO0907; THROMBOMODULIN.
DR SMART: SM00181; EGF_19.
DR PROSITE: PS000179; EGF_CA. 18.
DR PROSITE: PS00010; ASX_HYDROXYL. 14.
DR PROSITE: PS00022; EGF_1; UNKNOWN_1.
DR PROSITE: PS01186; EGF_2; 19.
DR PROSITE: PS01187; EGF_CA. 12.
DR PROSITE: PS00524; SOMATOMEDIN_B; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 937 AA; 101043 MW; 8c67830c8e391d07 CRC64;

```

Query Match 7.1%; Score 555.5; DB 5; Length 937;
 Best Local Similarity 20.4%; Pred. No. 7.3e-38;
 Matches 212; Conservative 88; Mismatches 259; Indels 481; Gaps 43;

```

QY 199 ACMIASLANAPAPGEVIMCA-----LGTASVMDGVNGTGCOCGLGFNG----- 244
DB 139 SCNNAAQPNNOQRNRYKIFCCPNPCNPSCGAYATCSNMHYRPVCTCNAGTTGNPOTG 198
QY 245 -TACETCTEGKYGIHCDAQSCVHGR---CSQGPLGD----- 277
DB 199 CTAIDPCNPSGGAHA--TCSSHYRPVCTCEAGYTGNGTCTDVNECLRPVNCGPASN 256
QY 278 -----GS--CDDCVGR--GVKCDMETTDNC--NGTCHTSAN--CLLDPDGASAK 321
DB 257 NKRCVNTASFRVCKNGRAQGSRC---VDINECLRPNICGINSNKRQVNRPGG--YRCV 312
QY 322 CAAGFNGNGTCAINACTSNGGCTKADCKR--TTPGNRVVCAGYTGSDIVCLEIN 379
DB 313 CRPGYRAQGSRCVDINECLRPN--VCGPASHNRCVNTPSFRVCNNGYRAQGSRCVDIN 371
QY 380 PCLENHGCDRNAEQTQTPNCAVNCNCLRPKYTGDKVCSLINCLINNGGCSPPACNCT 439
DB 372 EC--RSSPCGNNAQNCINT--PGSFTCKNGTGTNGRACIDRVNCAALIRPCSNACTNT 428
QY 440 EDDORICTCKPDYTGDIVCGRGSIYELPKNPSTSOYFQLOEHAVRELAGPGPFTVPAP 499
DB 429 -PGSFTCKCKPGYTGNGLVCR-----DINECLSRACG----- 460
QY 500 LSSSFNHEPRIKMDQGLMSQVLRHYVVGCCQLLDNLKVTTSATTLQGEPIVSISVD 559
DB 461 ----- 460

```

```

QY 560 TWFINNKAQVLSIIITNGVTHVIDKLLSPKNNLITPKDALGRVLQNLTTVAANHGYTK 619
DB 461 -----VNTN-----KRCINTPG----- 472
QY 620 FSKLIQDSGLSVITDSITPTVTFMPTDKALEALPPEQODFLFNQDNKDKLSYLFKHV 679
DB 473 ----- 472
QY 680 IRDSKALASDLPRASAMKTLQGSSELSVRCGTSIDIGELFLNEQMGREFIRGLFDGVAY 739
DB 473 -----SFR-----VCNGYR-----NQ 485
QY 740 GIDCLIMPFTLGRCDTFTTTPDIPEGGSCITTPKCPKLSKRGVAKKCIYNPFPRRNV 799
DB 486 GSRCDVIN-----EC-----RSSPCGNNAIRNTPGSF----- 513
QY 800 EGCQNICTYVIQTPRCCHGYFMPDOACGCGPDTPCNNNGMCRDLTPMGQCLCHFGNG 859
DB 514 -----TCRCNTSY-----TGNELICRDI-----NECEAH----- 537
QY 860 TACELCWHGRFGPDCOPRCSSEHGOCDEGITSGEICETGWTAAACDTPFAVAVCTPA 919
DB 538 -----PNCGENALCINGL--GSYRCICARGFSGPLC-IDFNCAAIRP 579
QY 920 GSVHATCTE--NNTCVNLNTEGDICTCTVVDPECKONNGGCAKVAKCSOKGTQVSCSK 976
DB 580 CSFNADCTFNTPGSFTQCKRPGYTGNGLVCRDINECSRPN--ACPRNQRCINTPGFNCVCA 638
QY 977 KGYKQGYSCIELDPDCAVNGGCHHATCRMTGPKKHCECKSHVGVGVCPEPQLPL 1036
DB 639 IGRKRYKNCVDINECR--ASRPRCDLNASCQNT--PSTFTCTCNTGTGNGLTC-----ADI 692
QY 1037 DKLQDNGCCHPDASCADLFQDPTVGVFHLRSPLQVYLTFEDKAKEACAKAETATYVN 1096
DB 693 NEC--NNPRACHQATCAN-----TP--GSY-----TCN 717
QY 1097 QLSYAKAKYHLCASAGLESGRVAYPTTYAS--QKGANVAYIVDYSRANKSEMDVFCY 1155
DB 718 -----CNRGYTGNGRVCFVNPSPSPSCSNAI----- 745
QY 1156 RMKDYN-----CTCKAGYVD 1171
DB 746 -CRAVNYRPVCTCKPQGTGN 764

```

```

RESULT 9
ID 09TVO2 PRELIMINARY: PRT: 1664 AA.
AC 09TVO2:
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Y64GI0A.7.
GN Y64GI0A.7.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Mortimore B.J.;
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
  investigating biology.";
RL Science 282:2012-2018(1998).
RN [3]
RP SEQUENCE FROM N.A.
RA Ainscough R.;
RL Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.

```

DR EMBL: AL117206; CAB60454.1; -
 DR EMBL: AL110498; CAB60454.1; JOINED.
 DR EMBL: AL110498; CAB57911.1; -
 DR EMBL: AL117206; CAB57911.1; JOINED.
 DR HSSP: P00736; 1A0Q.
 DR InterPro: IPR00152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR002049; laminin_EGF.
 DR Pfam: PF00008; EGF_25.
 DR PRINTS: PR00011; EGF_LAMININ.
 DR SMART: SM00179; EGF_CA; 4.
 DR SMART: SM00001; EGF_Like; 18.
 DR PROSITE: PS00010; ASX_HYDROXYL; 4.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_22.
 DR PROSITE: PS01186; EGF_2; 24.
 DR PROSITE: PS01187; EGF_CA; 3.
 DR Calcium-binding; EGF-like domain; glycoprotein; Hydroxylation; Repeat.
 KW SEQUENCE 1664 AA; 179279 MW; A69F093B4C705832 CRC64;

Query Match 6.88; Score 536; DB 5; Length 1664;
 Best Local Similarity 21.08; Pred. No. 8; Le-36;
 Matches 305; Conservative 142; Mismatches 451; Indels 556; Gaps 84;

23 IHYNLASAIESADAYTFVFNNEAIENYIR---EKATSLKEDILRHVHVGELKLNKD 78
 DB IFFELGTELEKD---FLPFPFHARKYLRFARFSRRGCS-KCCLLRVANCSADL----- 83
 79 LHNGMHRETMGLFSYLLAFRLNDOLYNEAPINYN-----VATDKGIHGLEK 128
 DB CHNG---CTCVPSBH-----NDNEQVCECPVGTGAKQCYDANECAANNNGGCEHEVN 133
 129 VLEIQRKRC-----DNNDIIVRGEGCKSQAP-----CLETKPLRETT 168
 DB 134 TIGTYRCRWGPFELSGDGNFTCSIDECVANSNGGSDRCVSPGFRCDPSDYLHADG 193
 169 KCIYSLIFEMKRSVFI-----GCOPQCVRTITRACMLASLAHNAKAPRGEVK 217
 DB 194 RTCCSGHFI--ENLLIKVITSCSTDNGGCEHC-----ENDSNGEFY 234
 218 MCALGTASVMDVNGTGTCCGGLGF---NCTAC---ETCTEGKYGJHCDQASCVRGC 270
 DB 235 -----RCRCRVGFELSENRKSCOPVDPCEFNKG--CQHHTNNHGR- 274
 271 SQGLGDSGCCDVGWGVKCDME--ITTDNC---NGTCHISANCLDPPDKASCACAG 325
 DB 275 -----AQCQCYEPGFH--LSYDRRCVDIDECANNNGCEHFCENV---KGTYRCRKEG 322
 326 FR--GNGTVC-TAINACETSNCGGCTKADCKRTTPGNFVYCKKAGY--TGDGIYCLEINP 380
 DB 323 YOLGRDRGTCEMLGCGOVNGGC--QHDICYDOPDGGHVCGRNCRYILLANDOKLCHIDINE 380
 381 CLENHGGCDRNAECTGTGPNQA---VCNCLP--RYTGDGKVCSLINCLTNNGGCSPFAF 435
 DB 381 CHENNG-----DCSOLCVNLASVCECKRPFRLMKRKTCEIDISECSNNGGE--QI 432
 436 CNYTEODORITCKPDY--TDDGIYCRGSITGELPKPNSTQYFFOLOEHAVER-LAGPG 492
 DB 433 CS-MQEGGYMCSCEPGEFELSEDHSC-----HDMNECLLNG 468
 493 PFTVFAPLSSFFNEPRIKMDQGLMSQVLRHVHVGCOQLLDNLKKTTSATTLQGPV 552
 DB 469 -----GCAQ-LCKNRKGRSCQCPAGT-I 490
 553 SISVSQDTVFINNEAKVLSDDISTNGV--IHVIDKLSPNLLITPKDALGVLDNLTT 610
 DB 491 LAHDEKSCVAASDADIFSDNIDYSKVPGLDSIDEVIS--SISYPADESPRL----- 543
 611 VAANHGTGTFKSLLODSLSLVTIDSIHTPTVFWPTTKALEALPPEQDPLFNODNKOK 670
 DB 544 VEGRRRHAKACVNFQGLTSLFESSEVRTD-----PSEKCPNGF----- 582

671 LKSYLKFHVIRDSKALASDLPRASAMKTLQSGELSRGCTGSD-IGELFLNEOMCRFIHR 729
 DB 583 -----FSTCOLSSDQNGCKGCMRSGSLSKDCFSGYTGEC--EOLCRNGW 631
 730 GLEFDVGAAYGIDCLMNPITLG--RCDFTTPTDIPGEC--GSC---IPTPCPLSKR--- 780
 DB 632 G-----VDCAHKSCSKLCPSTGSCRED-----PEKSDGCPDGFYGSQCNLKRMDC 681
 781 -----PKYKKCKIYNPLF-----RRNVCCONL---CTIV 809
 DB 682 PNGRCDPVGYCTCPDGLYGOSCEKPCPHFTFGKRCRPPCKARENSECEITGKC--- 738
 810 IOTPRCHGYFMPDQ-----ACPGG-----P 831
 DB 739 -----RCKPRTYGHNRKRCRSCGLGAGCAGMKSCDPAGIRCDPVYGDCTKKCPAGYQNL 794
 832 DTP-----CNNGRCRDLVYTPM----- 848
 DB 795 DQPCPAGYFGYDCEQKSCADAVASPHKSKVCHNYTGCTCLPCKTGPLCDGSCAPNTYGP 854
 849 -----GQCLCHTGFNGTAC-ELCHWGRGRPD-CPRSCSEHGQDEG 888
 DB 855 NCAHTCSCVNGAKCDESDGCHCTPGFYGATCSEVCPRGFRGIDCMOLCKONGAICD-- 912
 889 ITSGGECLENTGWTAAASCDPTAVFAVCTPA-----CSVHAFTTE-----NNTCVGNL 936
 DB 913 -TSNSCECAFGWSKCD-----KACAPGTFGKDSKCKDCADGMCDCPSDGEICTP 965
 937 NYEGDGI--TCTVDFCKONNGGCAKVAKCSOKGT---QVSCSKKQKGDYSCIEID 990
 DB 966 GKKGKHCDETCDSGLF-----GAGCKGICSONGATCDSVTGSCBRRPWRK--KCDR-- 1017
 991 PCADGVNG-GCH-----EAT--CR---MTP-----GKH 1014
 DB 1018 PCPGRFEGCGNAICDCTTNDTSMYNPVARCDHYTECHCPAGWGTGPDQOTSCPLGRH 1077
 1015 K-----CECKSHYVDGVDC--PEOLPLDRCLD-----DNGGCH 1047
 DB 1078 GEGCRHSCGCSNGASCDBRVTFPCDPSGFMKNCSECPREGIMSGNCKHCLCHNGGECN 1137
 1048 P--DASCADLYPDPTTVGFHLSPLOGYRLTFDKAKACAKAATATYATYNOISYAKA 1104
 DB 1138 KENGDCCECIDGM-----TGPSCEFLCPGQFGFRNCAORC---NCKNAGSDBRKT 1183
 1105 KYHLCSAGMLSEGRVATTTTASOKCGANYGYIDYGRASKEEMWDFYCYRMKDVNTC 1164
 DB 1184 GRCECLPGW--SGE-----HCEKSC--VSG--HYGAKCEET-----CEC 1216
 1165 KAGYVGDGFS--CS 1176
 DB 1217 ENGALCDPISGHCS 1230

RESULT 10
 09WU99 PRELIMINARY: PRT: 2906 AA.
 AC 09WU99;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE FIBRILLIN-2.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99350231; PubMed=10419698;
 RA Yang Q., Ota K., Tian Y., Kumar A., Wada J., Kashihara N., Wallner E.,
 RT "Cloning of rat fibrillin-2 cDNA and its role in branching
 morphogenesis of embryonic lung.";
 RL Dev. Biol. 212:229-242(1999).

DR EMBL: AF032672: AAC36152.1: -
 DR EMBL: AF032670: AAC36152.1: JOINED.
 DR EMBL: AF032671: AAC36152.1: JOINED.
 DR EMBL: AF032673: AAC36153.1: -
 DR HSSP: P00740: 1EDM.
 DR InterPro: IPR002110: ANK.
 DR InterPro: IPR000152: Asx_hydroxyl.
 DR InterPro: IPR000561: EGF-like.
 DR InterPro: IPR000742: EGF-2.
 DR InterPro: IPR001881: EGF-Ca.
 DR InterPro: IPR001438: EGF-II.
 DR InterPro: IPR000800: Notch.
 DR Pfam: PF00023: ank; 6.
 DR Pfam: PF00008: EGF; 36.
 DR Pfam: PF00066: notch; 3.
 DR PRINTS: PRO0010: EGFBL00.
 DR PRINTS: PRO1452: NOTCH.
 DR SMART: SM00248: ANK; 4.
 DR SMART: SM00179: EGF_CA; 24.
 DR SMART: SM00001: EGF_Like; 11.
 DR PROSITE: PS00088: ANK_REPEAT; 5.
 DR PROSITE: PS00297: ANK_REPEAT; 1.
 DR PROSITE: PS00010: ASX_HYDROXYL; 22.
 DR PROSITE: PS00022: EGF_1; UNKNOWN_34.
 DR PROSITE: PS01186: EGF_2; 28.
 DR PROSITE: PS01187: EGF_CA; 21.
 DR ANK repeat: Calcium-binding; EGF-like domain; Glycoprotein;
 KM Hydroxylation: Repeat.
 SQ SEQUENCE 2653 AA: 285928 MW: 66A2A058FE6C329 CRC64:

Query Match 6.6%; Score 519; DB 5; Length 2653;
 Best local Similarity 22.7%; Pred. No. 4,7e-34;
 Matches 298; Conservative 117; Mismatches 416; Indels 482; Gaps 84;

QY 137 CDNDNTIYRGEGKGSQAQRPLETRKRLRETRKCIYSIFEMK--RSVFICQPO-----190
 DB 203 CDNN-----PCKHNGTGC-LNT--HGSYQCMCPAGYTGKNCESKYPSPCQCN 248
 QY 191 ---CVRTIIRACWL-----ASLAHNAKPARGEVKKKALGTASVMDGVN-----231
 DB 249 GGTGRTSTGLYECKPREGYGKNCQENIDDCPGH--LCQMGGTCT-DGINSYHCACPPNY 305
 QY 232 -----GTGTGCGGLGFNGTACE-----248
 DB 306 TGENCKEDVDECAIRPSVCQNGATCTNSGYSICVNGMTGPCCSINIDDCIAAAGFYG 365
 QY 249 -----TCTEGKYGI--HCDQACS--C-VHRCSSQGL-GDGSQDCDVGMRGV 289
 DB 366 ATCIDGVGSFYCRCTPGKTLCLLHDADCTSNPCHADALCDTSPINGSYTCPCATGYKGV 425
 QY 290 KCDMEITTDGNCN--GTCHTSANCLLDPPDGKASCACAAGFNGNGTVC-TAINACTSNGGC 346
 DB 426 DCSEDI--DECDIGSPCEHNGVGVNTP-GSFRNCSSQGF--TGPRCTINIECESH--PC 478
 QY 347 STRADCKRTTPGRNVCCKAGYTG-----DGIVCLEINPCLLENHGGCDRNAECTQTGPNQ 401
 DB 479 QNNGSC-LDDPGTFRCVCMGFTGTGCEIDINEC-QSNPCL-NGICNDMI-----NG 528
 QY 402 AYCNCLEPKYTGDKVSL-INVCLT--NNGCCSPAFNGNYTEQDORICTCKPDYTGDG 456
 DB 529 FKSCSALGFNGSR--COINIDDCQSPCRNNGICRD-SINGY-----CQCPRGYT--G 577
 QY 457 IYRGSGIYGLRPNPSTISQVFFOLOEHAVALAGPGFT--VFAPLSSFNHPRIKMDQ 515
 DB 578 LSCSEINI-MNCNSNPC-----HRCICIDGNRFTCTVCPDGFGLYLCOTQDINECE- 625
 QY 516 OGIMSGQLRY--HYVGGQQLLDNLKVTTSATTLQGEPPVSISSQDVFVINNEAKVLSDD 573
 DB 626 ----SNPCQYGGGCHVDKRVGVMCHCLAGTS-----GKDCETINNE-----661
 QY 574 IISTNGVIHVIDKLSPKLLITPKDALGRVLONLTVAANHGYTKFSKLIODSGLLSVI 633

DB 662 -----CH-----SNPCNNKATCIDGI-----NKYTCQCVPGFT-----689
 QY 634 TDSIHTFYVYFWPTDKALEALPRPEQDFLEPNQDDKKDLKSLTKRHYIRDSKALASDLPRS 693
 DB 690 --GVHCEINI-----NECASNPANNGVCMDLVNGYKCECPGRFY---DPCLT-SDVDEC 738
 QY 694 ASWKTLOG-----SELVRCGTG-----SPIGELFLEQGCRTIHRGLDFDVGAV 739
 DB 739 ASNPCLNGKRCEDGIDINEFICHPRGYGGKRCENIDECSSNP--CQ--HGFVDELMNAF 794
 QY 740 GIDCLMNPVLGRCDFFTTFDIPGEC--GSCI-----FTPK-CPLKSKPK 782
 DB 795 KQCG--MPGYGLKCEINIDDCINNPCANGSTCIDKYNKYCVCKVYITQDDCSKLDPC 852
 QY 783 GVKR-----KCIYNP--LPR-----RVNRCQNL-----805
 DB 853 ATNRCRNDACCTPSNPLDPSCTCKLGYTGRCDEDDIDECKLSTPCRNNGATCHVAVPSYR 912
 QY 806 CTVVIGTPRCHGYFMPDCQACPGPDT-----PCNNGMCRD--LYTPMGQCLCHTFGN 858
 DB 913 CI-----CAKGYEGHDCAL--NTDDCAMFPQNGGTCLDGIQDYT-----CLCVDPGF 958
 QY 859 GTACEL-----CMHGRF-----GPDQCPRSCSEHGQC 885
 DB 959 GKHCETDINECLSMPCQNGATCRQYVNSYTCPCPLGFSGINCQTNDEDTCESSCMNGTIC 1018
 QY 886 DEGTGSGECLCEFTGMAASCDPTFAVAVCTPAGCSVHATCTENN--TCVCNINYGSG 942
 DB 1019 IDGI-NSYNCSCLPYGYGSSNQ--YKINKDSQPCQNGATCHEGDEYTCGSGYTGK- 1074
 QY 943 ITCT-VDPCKONNGGCAKAYAKCSOKTOYSCSKCKGYKGBGYSC-LEIPCAD-----G 995
 DB 1075 -QCTDIYDWCKTS--PEENCATCTQYKNQFSCKRAPWTGK--LCDYEMVSCSIALRK 1129
 QY 996 V-----NGCCHENATCRMTGPKHCKECKSHYVG-----1024
 DB 1130 VSLEQLCNNGTCKEHNMI-----HRCYKQGYTGYSQDEINECESQPLANGTCRDL 1182
 QY 1025 -----DGVCEPEPDPLDRC-----LDNCGCHPDASCADLYFDOTTYGVNHLR 1068
 DB 1183 IGSYACVCRKFGQONE--LNLIDCSPNPCQNGGTCH-----DLP-----VNTFSCS 1227
 QY 1069 SPLGOYKLPFDKAKEACAKEMATATATYNQLSYAQKAYHLCASGLBSGRAVYPTTASQ 1128
 DB 1228 CPCTACILICEVNEIDCKRSGC-----HNGTCTIDRVGGFEC-----ACPPGVGS 1273
 QY 1129 KCGANV-----VGIVDYSRANKSEMDVFCYRM-KDVNCTCKAGYVG 1170
 DB 1274 RCEGDINECLSNPCSNAGTLD-----CVQLVNNYHNCNCKKPGYMG 1312

RESULT 12
 Q9W0H8 PRELIMINARY; PRT: 2872 AA.
 ID Q9W0H8;
 AC Q9W0H8;
 DT 01-NOV-1999 (Tremblrel. 12, Created)
 DT 01-NOV-1999 (Tremblrel. 12, last sequence update)
 DT 01-DEC-2001 (Tremblrel. 19, last annotation update)
 DE FLBRILIN-1.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90032689; PubMed=9815129;
 RA Kanwar Y.S., Ota K., Yang Q., Kumar A., Wada J., Kashihara N.,
 RT "Isolation of rat fibrillin-1 cDNA and its relevance in metaphoric
 development.";
 RL Am. J. Physiol. 275:F710-F723(1998).
 RN [2]

RP SEQUENCE FROM N.A.
RA Kanwar Y.S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF135059; AAD34438.1; -
DR HSSP: P35555; 1AP1
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR002212; TB.
DR InterPro: IPR000823; ZnF-C2H2.
DR Pfam: PF00008; EGF_45.
DR Pfam: PF00683; TB; 9.
DR SMART: SM00179; EGF_CA; 41.
DR SMART: SM00001; EGF_Like; 5.
DR PROSITE: PS00010; ASX_HYDROXYL; 42.
DR PROSITE: PS00022; EGF_1; UNKNOWN_2.
DR PROSITE: PS01186; EGF_2; 38.
DR PROSITE: PS01187; EGF_CA; 41.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_1.
DR Calcium-binding; EGF-like domain; Glycoprotein; Hydroxylation; Repeat.
KW SEQUENCE 2872 AA; 312069 MW; 0CAFE3B87A80280 CRC64;

Query Match 6.48; Score 499.5; DB 11; Length 2872;
Best Local Similarity 21.88; Pred. No. 2,4e-32;
Matches 251; Conservative 111; Mismatches 322; Indels 469; Gaps 67;

QY 189 POCRTITITRACWLAS-LAHNAK--PAPG-----EYKMCALGTASV 226
DB 443 PKYLPENVTDYCOLVRYLCOGRICPTIPGTYRCECNKGFOLDIRGECDVECEKNPCTG 502
QY 227 WDGVNCTG--TCQCGLGFNGTACTCTEGKYGICDO-ACSCVHGRSGQPLGDS--CD 281
DB 503 GECINNGSYTCHCRAGYQSTL--TRTECRDIDECLONGRCINNGRCIN--TDGSPFCV 557
QY 282 CDVGWR---GVKCDMETTDNC-----NGTCHTSANCLDPPDGKASCACAFR--G 328
DB 558 CMAGFHVTRDGKNCE---DMDECSIRNCLNGMC-----INEDGSEKCTCKPGFOLAS 607
QY 329 NGYVCTAIIACET---SNGGC-----KRTTPGNRV-- 361
DB 608 DGRCKDINECEETPGICMGRCAVNTDGSYRCEPGLAVGLDGRVCVDTMHRSTCYGYR 667
QY 347 -----STRADC-----KRTTPGNRV-- 361
DB 668 RGCCVCKPLFGAVTRKECCASTEFYAFGPACOPCPAONSAYEQALCSSGSPRTAGSDINE 727
QY 362 -----CVCKAGYTDG--GIVCLEINPCLNHCGRNAEACTYTG 398
DB 728 CALDPDICPNIGICENLRGYKICICNSGEVDITGRKNCVDINECVLNSLLCD--NGCART-- 785
QY 399 PNOAVCNCLPK---YTGDGKVCSLINCLTN--NGGC--SPFAFCVNTBODQRICTCKP 450
DB 786 PGSEVVCIC-FKGFYRKDLKTCEDECESSPCINGVCKNSPSSF-----ICECSP 835
QY 451 DYTGD--GIVCGRGSIYELPKNPSTSQYFQLOEHAVERELAGPFTVFAPLSSFNHEP 508
DB 836 ESTLDPTKICIEITIKGT----- 853
QY 509 RIKMDMOGLMSQVLRHVHVCQQLLD--NLKVTTSATTLQ-----GEPSVI-- 554
DB 854 -----CMQTVIDGRCEININCATLKSECCSSIGAMGSPCTIQ 892
QY 555 -----SVSODTFVINEAKVLSDDIISTNGVI--HVIDKLSPNLLITP 597
DB 893 VDPICGKGYRIKGTGCEDI---NECEVFRG--VCKNGICLVNRSKSCCPGSMIL-- 944
QY 598 KDALGRVQLNLTVAANHGYTKFSKLLQDSGLSVI---TDSIHTPTVTVFWPTD----- 648
DB 945 -DATGRICLDRL-----ETCFLLKYDDECELTPIAGNHRMDACSSVGAAMGEEDEEC 997
QY 649 -----KALEALPREQDFLRNODKKDLKSTYLKFNHVRDSALASDLPRASMKWTLOGSE 703

DB 998 PLRNSREYELCPRGPGA---TKDITNGKPFKDIINECKMIDSLCTHGKCRMTI--GS 1051
QY 704 LSVRCGTSGDIDGELFLINOMCRFIHRGLLPDVAAYGIDCLLMNPDTGRCDDTFPTDIP 763
DB 1052 FKCRDSDGFALDS---ERRNCT-----DIDECRISPDLCR----- 1084
QY 764 GECCSCIFTP-----KCPYKSKPRG--VKKKCIYNPLPFRNRVBCQNLCTVVIQTPRCC 816
DB 1085 ---GQCVNTPGDEFECKCD-EEYESGFMKMKCM-----DIDECQ-----RDPILLC 1125
QY 817 HGTFMPDCA---CPRGPD-TRCNNGRCRDL-----YTPMGCL-----C 853
DB 1126 RGGICHTNTEGSRCEBPSGHOLSP--NISACIDINECELSANLCPHRCVNLIGKYECAC 1183
QY 854 HTGFNGT-----ACE-LCWV--GRECPDOP-----RSCSEHQC 885
DB 1184 NPGYHPTHDLFCVYDIDECSTIMNGCEFCFINSDSGYSQCPGFALMPDORSCDIDDEC 1243
QY 886 DE-----GITSGECLCETGWTAA-----SCDPTPAVFAVCTPACSVHA 924
DB 1244 EDNPNICDGGGCTNLPGEYRCLCYDGFMASEDMTKCYDVNCCD-----LNPNICLSG 1295
QY 925 TCTENN-----TCVCNLANEG-DGIT-CTVDFPCQONNGCKAKVAKCSQGTQVSCSKKG 978
DB 1296 TC-ENTKGSFICHCDMGSGKKGKTGCTDINECEIGAHNCRHAAVCTFTAGSFRCSGSPG 1354
QY 979 YKGGYSCIEIDPCADGVNGGSCHEATCRMTGPKHKCEKSHVVGDDVDCPEQPLPDR 1038
DB 1335 WIGGIKCTDIDECSSNGTH-MCSQHADCKNT-MGSIYRCLCKDGTGGDFTC---TDLDE 1408
QY 1039 C-----LQDNGQC 1046
DB 1409 CSENLNLSGNGQC 1421

RESULT 13
ID 035516 PRELIMINARY; PRT; 2470 AA.
AC 035516;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CELL SURFACE PROTEIN.
GN NOTCH2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57B/6; TISSUE=THYMUS;
RX MEDLINE=93178563; PubMed=8440332;
RA Lardelli M., Lendahl U.,
RT "Notch A and Notch B - two mouse Notch homologues coexpressed in a
RT wide variety of tissues."
RL Exp. Cell Res. 204:364-372(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57B/6; TISSUE=THYMUS;
RA Hamada Y., Higuchi M., Tsujimoto Y.,
RT "Complete amino acid sequence and multiform transcripts encoded by a
RT single copy of mouse Notch2 gene."
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
DR EMBL: D32210; BAA22094.1; -
DR HSSP: P16109; 1FSB.
DR MGD: MGI:97364; Notch2.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF_Like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_CA.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.

DR Pfam; PF00023; ank; 6.
DR Pfam; PF00008; EGF; 35.
DR Pfam; PF00066; notch; 2.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 4.
DR SMART; SM00179; EGF_CA; 22.
DR SMART; SM00001; EGF_Like; 12.
DR PROSITE; PS50088; ANK_REPEAT; 4.
DR PROSITE; PS50097; ANK_REPEAT; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 22.
DR PROSITE; PS00022; EGF_1; UNKNOWN_33.
DR PROSITE; PS01186; EGF_2; 27.
DR PROSITE; PS01187; EGF_CA; 22.
KW ANK repeat; Calcium-binding; EGF-like domain; Glycoprotein;
KW Hydroxylation; Repeat
SEQUENCE 2470 AA; 265325 MW; B55A31B35242716D CRC64;

Query Match 6.2%; Score 489; DB 11; Length 2470;
Best Local Similarity 22.7%; Pred. No. 1,5e-31;
Matches 282; Conservative 119; Mismatches 445; Indels 398; Gaps 78;

QY 132 IQKNRCNNMT-----IIVGEGG-KCSQAAPCPLETKPLRETRKCIYSIFMCK 180
DB 302 LQPNACMGCTCTNRNGGVCVNGWSQDCESENI-----DCAYA----- 343
QY 181 RSVFIGCQPCVRTIITRACWLASLAHNAKPARGEVKKMCALGTASY-----WDG 229
DB 344 -----SCPPG--STCIDRASFSCLPBEGKAG-----LCHLDACISNCHKALCDTNP 392
QY 230 VNGTGCQCGLGFCNCTGACPTCEGKYGIHQDQAS--CVH--GRSGGPLDGS--CDQCV 284
DB 393 LNGYITCTPCQGYKACDTEVD-----ECAMANSNPECHACKVNA--TDAPHCBLK 444
QY 285 GWRGVKCDMEITTDNCG-TCHTSANCLLDPDGKASCACAGFRNGTVC-TAINACET- 341
DB 445 GYAGRCRMDI--NECHSDPCQNDATG-LDKIGFTCLCMPEFK--GVNCELEVBCGN 499
QY 342 ---SNGGSTKADCKRTTPGRNV-CVCAAGTYGDGIVC-LEINPCLBNHGCGDRNAECTO 396
DB 500 PCVNNGGQCDKV-----NRFQCLCPGGTGP--VCOIDIDDC--SSTPLMNAKCID 547
QY 397 TGPNAVNCNCLPYTGDDGVCSLIWCLTNNGCSPPAFNCNTG---QDQITCCKPY 452
DB 548 -HPNGYECOCATGFTG-----ILCDENIDNCPDP-CHHGQCCODGIDSTICTCNPEY 597
QY 453 TGDGIVCRGSIYGLPKNPSTSQYFFLOEHAHVRELAGGPTVPAPLSSFNHEPRID 512
DB 598 M--GAICSDQI-DECYSSPC-----LNDGRCIDLVN-GYOCNQOPGTSGLNCETNFD 646
QY 513 WDOGGLMSQVL-----RHHVGCQQLLDNLKVTTSATTLGCEPVSISVSQDTVEINMEA 567
DB 647 CASNPCHMGVYCDGINRYSCV-----CSPGFTGQRNIDIDE---CASNPC 689
QY 568 KYLSDDIISTNCVIVHIDKLSPKMLITTPKALGRVIONLTVAANHYTEFSKLIDS 627
DB 650 RKGATCINDVNGRCICTCEGPHHPSCYSOVNCLSNPCIGHCTGGLSY---KCLDA 745
QY 628 GLLSYITDSIHPTVVFVPTDKALEALPPEODLFFNODNKDKLSYLKFHVRISKALA 687
DB 746 GWVGYNCE-----VDKNECLSNPCQNGTCN--NLVNGY----- 777
QY 668 SDLPASAKMTIQSELSYKCGTSGDIGELINQMCRPIHRLGLFDVGAVAGIDCLLMN 747
DB 778 ---RCTCKKGGKYNCOV-----NIDECA SNP--C--LNGGTCFDDVSGYTCMCL- 821
QY 748 PTLGRCQPTTFDIPREG-----SCITFP-----KCPKX-----SKP-- 781
DB 822 PYTGNCOITVLAFCSPNCEANAAYCKEADNPESFSCCLAPGCKGRCYDVDECI SKPM 881
QY 782 -KGV-----KKCIYNPLPFRNVEGCON--LCTVVIQTPR 814

DB 882 ANGVCHNTGGSYVCECPFCFSMGDCEDINDCLAMP-----CQNGSCVDHVNIFS 932
QY 815 C-CHGYFMPD-CQA-----CPGSPDTPCNNRGMCRLLYTPMGQCLCHTGFNGTACE----- 863
DB 933 COCHGFFIGDKCQDTMNEC---LSEPCKNKGTCSD-YVNSYCTCPAGHGVCHENNIDE 988
QY 864 -----LCWHGRFCP-----DCQPSGSEHGQCDGITGSGEC 895
DB 988 CTRESSCFNGCTGVDGINSFCLCPYGFPGFCLHIDNECSSNPCLNAGTCVGL-GTYRC 1047
QY 896 LCETGWTAAACDTPPAVEAVCTPA-CVSHATCTENNT---CVCNINYESDGIITCTYVDF- 950
DB 1048 ICPLEGTGKNQCT---LVNLCSRSPCKMKKGTGVQKARPHCLCPGWM--DGAYCDVLNVS 1102
QY 951 CK-----QNNGGCAKAKKSQKGTQVSCSKKRYKKGDDGYSCLIEIDPCADGV 996
DB 1103 CKAALQKGVPEHLCOHSGTICINA-----GNTHHQCPILAGTYG-SYCEEQLDCECA-- 1152
QY 997 NGGCHHATCRMTGPGKHKCECKSHYVGDPCEPEOLPLDRCLQDNGCHPDASCADLY 1056
DB 1153 SNPCQHGATCN-DFIGRYRCEGVPGY--QGVNCEYE---VDEC--QNOPCQNGTCTIDL- 1203
QY 1057 FQDTTGVGFHLRSLPGQYKLPFDKAKACAKAEATATATYMNLSYAQAKYHLSAGMLES 1116
DB 1204 -----VNHKSCSPGTRCLICEENIDCA-----GSPHCLNGGQCV 1241
QY 1117 GVAVPTT-----YASQKCANVGVIVDYSRANKSEMDVFCYRMK-DYNCTCKAGTYG- 1170
DB 1242 RIGGYTCRCLPFGAGEREGDINECL---SNPCSE-GLSDCVOLKNVNCICRSATGCR 1297
QY 1171 -----DGFSC-----SGNLQ 1181
DB 1298 HCTELDVCPOKPCPLNGGTCVAVSNMPPGFCRCPPGSGARLQ 1341

RESULT 14
09GPA5 PRELIMINARY; PRT; 2524 AA.
AC 09GPA5.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PUATIVE NOTCH RECEPTOR PROTEIN.
GN NOTCH
OS Brachyostoma floridae (Florida lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchyostomidae;
OC Branchyostoma.
OX NCBI_TaxID=7739;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LARVA;
RA Holland L.Z., Burgtorf C., Holland N.D., Lehrach H., Tamme R.,
RA Abl-Rached L., Pontarotti P., Lardelli M.;
RT "Cloning and developmental expression of the amphioxus homologue of
RT Notch (Amphinoctch): evolutionary conservation of multiple expression
RT domains in amphioxus and vertebrates";
RL Summary (Apr-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y12539; CAC19873.1; -;
DR InterPro: IPR002110; ANK.
DR HSSP; P00740; IEDM.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR002049; Laminin_EGF.
DR InterPro: IPR000800; Notch.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00066; notch; 3.
DR PRINTS; PR00010; EGFBLD.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR01452; NOTCH.

DR SMART; SM00248; ANK; 6.
 DR SMART; SM00181; EGF; 37.
 DR SMART; SM00179; EGF_CA; 34.
 DR SMART; SM00001; EGF_like; 13.
 DR SMART; SM00004; NL; 3.
 DR PROSITE; PS50088; ANK_REPEAT; 4.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS00010; ASX_HYDROXYL; 1.
 DR ANK repeat; EGF-like domain; Glycoprotein; Hydroxylation; Receptor;
 Repeat.
 KW SEQUENCE 2524 AA; 270969 MW; C2CA57E306D23BC9 CRC64;

Query Match 6.2%; Score 486.5; DB 5; Length 2524;
 Best Local Similarity 22.1%; Pred. No. 2,5e-31;
 Matches 271; Conservative 107; Mismatches 413; Indels 437; Gaps 72;

QY 236 COCGIGFNGTACETCTEGKYYGHCQACSCVHG-----RCSGG 273
 DB 233 CHMSGFTGDDDEVVDDCVHLCENGACVGVNEYTCTCPSONAGRYCNEDVDECMQS 292
 QY 274 P-----LGDGSCDCVGNRGVYKCDMEITTDNC-NGTCHTSANCLDPPDGKASC 320
 DB 293 PICTLNGCTCHNTVGGYSCVYNGMIGDCCSENF--DDCASAACEFGATC-HDRVGFEMC 349
 QY 321 KCAAFERNGVCTAIIACETSNCGCSTKADCKRTTP--GNRVYCKAGYTGDIYCL-E 377
 DB 350 ECAPG--KTGLLCHLDADACESS--PCNEGALCD-TNPVNGOPICPCPDGY--EGQLCMQD 402
 QY 378 IMPCLNENGGCRNAECTGTGPNQAVCNCLPKRYTGDGKVCSL-INVCLTNNGSCSPFAC 436
 DB 403 IDECALGENPCHDEGCNNV-PGSFTCTCTDGTGDR--CEVINMECASN--PCONOGTC 457
 QY 437 NYTEDDORICTCKPYTGDIYGRSISYELPKNP-----STSOYFQLO----- 481
 DB 458 -IDDIGEFACAMPFAGD--LCETDV-DECASSPCLNCLCDGINKYCECDPREGTT 513
 QY 482 -EHAVERELAGPPTFYFAPFLSSSFNHEPRIKMDQGLMSQVLRVHVYVCGQLLDNLKY 540
 DB 514 CENNINECAN-GPCRNGAHC-----DLVT 537
 QY 541 TTSATTLQG-----EPVISYSQDTVFINNEAKVLSDDIISTNGVT--HYDKL- 587
 DB 538 TYACTCLEGTGDEINIDQSNOCCHGTGVGASFTGCSEPCYGNPLCESPVEDC 597
 QY 588 -----LSPKLLITPKDALGRVL-----QULITNVAANH 615
 DB 598 SPCQNGTCELDVNGYRNCNLAGTSNGCEYNODCTGNLCYHGVCGDGLNDYTCQCRG 657
 QY 616 GYTK-----FSKLIQDSGLSVITDSIHTPVYFWPTDKALEALPPEQDQFLNQ 665
 DB 658 GYEANNCEREIDECASSPCHNGI-----CHDLVNAE--SECPRGYHDLQCYEN 705
 QY 666 DN-----KDKLSKY-LKFEIVIRDSKALASDLPRSS----- 695
 DB 706 VNECESSPCAHGTCDGINDYCTCENGEGKNCVDNIDECSNPNQHEGQDDGIGRYE 765
 QY 696 KRTLOGSELVRCGTSGDIGELFLNEQMCRTIHRGLLPVGAAYGIDCLMNPITLGRCD 755
 DB 766 CCGLCGYE-GVMCDINTD--ECASNP--CQ--NGRCILDGVMNNYVDDCL--PFTVGTNQ 816
 QY 756 TPTTIDIGEC---GSCI-----FTPKPLKSKPKGVK-----KKCIYNPLPFR 797
 DB 817 TELAPCRPMPENLACIPISADYQFTFCNC-----ADFEGETCADDINECSNPNC--K 868
 QY 798 NVEGCONLCTVVIQTPRCCHGYFMPDC-----QACPGSPD---TPCNRNRCMRD--L 844
 DB 869 NCAPCNL-----EGDFRCDCLMFAGELCSVINIDODDPCCNGSGTCDNGINS 917
 QY 845 YTPMGQCLCHTGENGTACEL-----CMHGRFGPDCQPR- 877
 DB 918 YT-----CSCMPGFGTNCEDIEDICTSNPCQNGGOCIDAVNGYACDCVVGFTGNTQTKN 973

QY 878 -----SCSEHGOCDEGTTGSGECLCTGTWTAACDTPTAFAVACTPACSVHATCTENN- 930
 DB 974 DDCSTSSCSFGSTCIDG-NTFTCHCPSGFTGSNQ--HEINECSNPNQNGATCVDQTC 1030
 QY 931 --TCVGNLNEYEDGTTG-TVPDPCKN---NGS----- 957
 DB 1031 YFSCICTGYE--GVYCSQKDLCADDDPCRNSTGTCSGDRYECLCEDEMTGLIDMTKV 1088
 QY 958 -CAKAK-----CSQKGTQVSCSKKKYKGDGYSCIEIDPCADGVNGGCH 1001
 DB 1089 SCMAAASGRNNSLANLNCNGGTCVDTGNSHNCCNCAAGYNG-SYCSSEIDECA---SPQ 1144
 QY 1002 EHATCRMTGPKHKECKESHVYGVGVDECEPEQLPDRCLQDNGGCHPDAACADLYFDOT 1061
 DB 1145 NCAEGR-DGLGYTTCACRPGY--QGVNCEOE--INECT--SNPCQNGGTCTIDM----- 1190
 QY 1062 VGVFHLRSLGQYKTLFPDKAKENAKKATITATYNOLSYAKAKYH--LCSAGNLESGRV 1119
 DB 1191 VNEYRSCPPGTGGLLCEINNDNCFAGAC-----YHDTGVVDGIEFTCR 1235
 QY 1120 APTTYASQKGCANY-----VGIVDYSRANKSEMDVFCYRMK-DVNYCTCKAGY 1168
 DB 1236 CRP-GYVGRCEGVDNECLSNPCDAEGTLD-----CVQLENDYSCDCKRXY 1280
 QY 1169 VG-----DGESC--SGN 1178
 DB 1281 TGRRCERTVDSCEPDLCLNGACASQSGN 1308

RESULT 15

Q9XWD6 PRELIMINARY; PRT; 1111 AA.

AC Q9XWD6;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE Y47H9C.4 PROTEIN (CED-1).
 GN Y47H9C.4 OR CED-1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Harris B.;
 RL Submitted (OCR-1998) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
 RA Gardner A., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gartner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., Murray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN (3)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21097720; PubMed=11163239;
 RA Zhou Z., Hartwig E., Horvitz H.R.;
 RT "CED-1 is a Transmembrane Receptor that Mediates Cell Corpse
 Engulfment in C. elegans.";
 RL Cell 104:43-56(2001).
 DR EMBL; AL032657; CAA21739.1;
 DR EMBL; AF332568; AAG60061.1;
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.

DR InterPro: IPR002049; Laminin_EGF.
DR Pfam: PF00008; EGF; 9.
DR PRINTS: PR00011; EGF_LAMININ.
DR SMART: SM00180; EGF_Lam; 6.
DR SMART: SM00001; EGF_Like; 5.
DR SMART: SM00261; FU; 2.
DR PROSITE: PS00022; EGF_1; UNKNOWN_15.
DR PROSITE: PS01086; EGF_2; 11.
KW EGF-like domain; Glycoprotein.
SEQUENCE 1111 AA; 118803 MW; A39F374C008F9874 CRC64;

Query Match 6.1%; Score 480.5; DB 5; length 1111;
Best local similarity 22.2%; Pred. No. 2.2e-31;
Matches 251; Conservative 98; Mismatches 341; Indels 441; Gaps 70;

QY 115 NVADTKGYI--HGLEKVLTI-----OKNRCDN-----NDTIYRGEGCKGCSQAQPCPLETK 163
DB 43 HVCYKTIIVDDYELKFKVIHTVYVNDTEQCLNPLTGFQCIKERGO--KASYROL---VK 97
QY 164 PLRETRKCIYTIYFMGKRSVFICQPOCV--RTIITRACWLASLAHNKAPAEVKKMCA- 220
DB 98 KEKYVKQCCDGYVQTKRHHFCLIPDCNPCKKKGKIEPGKC-----ECDDGYGG--KYCAS 149
QY 221 -----LGTASVMDGVNGT-----GTCCGLGFNGTACE--TCTEGKYGIHDOACSC 265
DB 150 SCSTVGTWGLGCKSCDCCEGANCDELGTCTCTSGFQGERCEKPCPDNKNWGPNCVAKSCPC 209
QY 266 VH-GRCSSGPILDGSCDCDVGWR-----GYKCDMEITTDNCNGTCHTSANCL 311
DB 210 QNGCKCKN---EGKCVCSDGWGEFCNLKCEKKGAECKFE-----CN--QNGATC- 257
QY 312 LDPDCKASCK-----CAAGFRNGTV--CTAIN--ACETISNG----- 344
DB 258 DTTNGKCIKCSGYHAGLCBENECVGFPGSGCTQKCDCLNMONCDSSSECKCIQWTGKHG 317
QY 345 --GCST--KADCKR--TTPG-----NRVCYKAGYTDGIVCLEINPCLEN 384
DB 318 DIGSGRGRGLQCKQNCQCPGLEFSDSNASCDAKTGQCCCEGKYGP--KODERKCDARO 375
QY 385 HGG-CDRNAECTQ-----TGPNAVCNCLPKYTGDKVCSLINVCLTNNG--CSPPAFC 436
DB 376 YGADCSKTCTCVRENTLMCAPNTGFCRCKPGFYGDN--CEL--ACSKDSYGFNCEKQAMC 431
QY 437 NTEGDDO-----KICTCKPDYTGDDGIVCRGSTYGLPKPMPSTSYQFFQJQEHAVRELAP 491
DB 432 DMNHASECNPETGSCVCKPGRGTG-----KNCSE-----P 460
QY 492 GPFTVFAPLSSSFNHEPRIKDMDOGLMSQVLRHVVGCCQQLLDNLKVTTSATTLQGP 551
DB 461 CPLDFYGP--NCAHQCC--NQRG-----VGCD----- 484
QY 552 VSIYSODTFLINFEAKVLSDDIISTNGVHVIVDKLLSPKULLITPKDALGRVQLNLTIV 611
DB 485 -----GADGKCCODRGWTH 499
QY 612 AANHGYTKFSKLIQDSGLLSVTDSITHTPYVFWPTDKALEALPEEQDFLFNODNKDL 671
DB 500 RCEH-----HCPADTF----- 510
QY 672 KSYLKFVIRDSKALASDLPRASAWKTLQGSLSVRCGTGSDIG--ELFLNEQMRIFHRG 730
DB 511 -----GANCKEKRCKPKGICGDPITGECTCPAGLOG 541
QY 731 LLEFDVGV--AVGICLLMNPILGRCOTFTTFDIPGECGSC---ITPKCPILK--SKPKG 783
DB 542 ANCDIGCEPGSYGPGCKLHCKCVNGKCDKET-----GEC--TCQPGFSGSDCTGSKGR- 594
QY 784 VKKCIYVNLPRRVNVEGQNL-----TVVLOTPRCCGCHGYMPOQACPGGP-----DT 833
DB 595 -----YG-----ESCELSGPCSDASCSKQTKGC-----LCPLGTKGVSQDQ 630
QY 834 PC--NNRG-MCRDLYTPM-----GOCL-CHTGFGNGTACE-LCMHGRFGPDQP--RS 878

DB 631 KCDPMTFGFLCQETVTPSPSCASTDPKNGVCLSCPPGSSGICHEHNCPAGSYGDGCOQVCS 690
QY 879 CSEHOCDEBITGSGECCICEGTWTAASCD--TPTAVF-----AVCTPACSVHATCTE--NMT 931
DB 691 CADGHGCD--PTTGECICEPGYHGNKTCSEKCPDKRYGYGCALDCPKCASGSTCHINGL 747
QY 932 CVCNINYESGDGITCTVVDFFCGKONNGGCAKAVAKCSQKGTQVS-----CSCKKGYKGDGYSC 986
DB 748 CICIPAGLE--GALCTRPCSAGFGWNGCGRVCRCTSEYKCKCNMQTEGCSCPAGFGD--RC 803
QY 987 IEIDPCADGVNG--GCHEHATCRMTGPG-----KHKCECKSHYVD-----GYDC 1029
DB 804 DK--PCEDGYVGPDCIKKCKCGTATSSCNRVSGACHCHPGFTGEFCHALCESTFGLKC 861
QY 1030 EPEQLPLDRCLDNGOCHPDASACADLYPDOTTGVYFHL-RSLGQYKLTFD 1079
DB 862 SKE-----C--FKDCCGDDYECDDAAIGCCHVDQMSGKRAKOFFE 898

Search completed: June 17, 2002, 12:35:05
Job time: 321 sec
